PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/57, 15/62, 9/64, 15/11, 5/10, A61K 38/48, C07K 16/40, A61K 39/395, A01K 67/027, C12N 15/00, C12Q 1/37, 1/68, G01N 33/574, A61K 31/70, 48/00

A3

(11) International Publication Number: **WO 98/11238**

(43) International Publication Date: 19 March 1998 (19.03.98)

(21) International Application Number: PCT/US97/16175

(22) International Filing Date: 11 September 1997 (11.09.97)

(30) Priority Data:
60/025,301 13 September 1996 (13.09.96) US

(71) Applicant: DANA-FARBER CANCER INSTITUTE [US/US]; 44 Binney Street, Boston, MA 02115 (US).

(72) Inventors: ANISOWICZ, Anthony; 50 Upham Street, West Newton, MA 02165 (US). SAGER, Ruth; 20 Codman Road, Brookline, MA 02146 (US). SOTIROPOULOU, Georgia; 1 Feidlou, Kastellokambos, Rion, GR-261 10 Patras (GR).

(74) Agents: MANDRAGOURAS, Amy, E. et al.; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report:
4 June 1998 (04.06.98)

(54) Title: PROTEASE M, A NOVEL SERINE PROTEASE

(57) Abstract

Isolated nucleic acid molecules encoding a novel serine protease, Protease M, is disclosed. Protease M is downregulated in metastatic mammary epithelial tumor cells, as well as other tumor cells, and is upregulated in senescent cells. In addition to isolated nucleic acid molecules, the invention provides antisense nucleic acid molecules, recombinant expression vectors containing a nucleic acid molecule of the invention, host cells into which the expression vectors have been introduced and non-human transgenic animals in which a Protease M gene has been introduced or disrupted. The invention further provides isolated Protease M proteins, fusion proteins, antigenic peptides and anti-Protease M antibodies. Diagnostic assays, drug screening assays, and therapeutic methods utilizing compositions of the invention are also provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		

PROTEASE M, A NOVEL SERINE PROTEASE

Background of the Invention

Under normal growth conditions, cell proliferation is tightly regulated in response to diverse intra-and extracellular signals. This is achieved by a complex network of protooncogenes and tumor-suppresser genes that are components of various signal transduction pathways. Activation of a protooncogene(s) and/or a loss of a tumor suppresser gene(s) can lead to the unregulated activity of the cell cycle machinery. Tumor suppresser genes can be divided into two classes. Class I, in which a loss of function results from a mutation or deletion and class II, in which a loss of function results from a regulatory block to expression (Lee et al. 1991. *Proc. Natl. Acad. Sci.* 88:2825). Thus, both activation and loss of genes can lead to unregulated cell proliferation and to the accumulation of genetic errors which ultimately will result in the development of cancer (Pardee, *Science* 246:603-608, 1989).

Malignancy is defined as neoplastic growth that tends to metastasize (Stetler-Stevenson et al. 1993 *Annu. Rev. Cell Biol.* 9:541). Metastasis is a multistage process involving numerous aberrant functions of the tumor cell. These aberrant functions include tumor angiogenesis, attachment, adhesion to the vascular basement membrane, local proteolysis, degradation of extracellular matrix components, migration through the vasculature, invasion of the basement membrane, and proliferation at secondary sites (Poste, G. and Fidler, I.J. (1980) *Nature* 283:139-146; Liotta, L.A. et al. (1991) *Cell* 64:327-336). Therefore, accumulative changes in the expression of multiple genes probably occur before tumor cells acquire the phenotype that enables them to metastasize. The identification of genes involved in the development of the metastatic phenotype is essential for an understanding of the molecular mechanisms underlying metastasis and for the design of novel therapies designed to arrest progression of a primary tumor.

Increased proteolytic potential is one documented feature of the metastatic phenotype. This increased potential is thought to result from the combined aberrant regulation of proteolytic enzymes (e.g., metalloproteinases and serine, cysteine and aspartyl proteinases) and their endogenous inhibitors (for a review, see e.g., Sloane, B.F. and Honn, K.V. (1984) *Cancer Metastasis Rev.* 3:249-263). For example, increased activity of serine proteases has been implicated in metastasis (Testa et al., Cancer metastasis Rev. 9:353, 1990; Dano et al., Adv. Cancer Res. 44:139, 1985; Ossowsky, Cancer Res. 52:6754, 1992; Sumiyoshi, Int. J. Cancer 50:345, 1992; Duffy et al., Cancer Res. 50:6827, 1992; and Meissauer et al., Exp. Cell Res. 192:453, 1991. In

addition, other proteases have been shown to be involved in augmenting tumor cell invasion, such as metalloproteases (DeClerck et al., Cancer Res 52:701, 1992), Wolf et al., Proc. Natl. Acad. Sci. USA 90:1843, 1993; and Sato et al., Oncogene 7:77, 1992) and cathepsins Rochefort et al., Cancer Metast. Rev. 9:321, 1990; and Kobayashi et al., Cancer Res. 52:3610, 1992).

Serine proteases are protein cleaving enzymes, which contain a serine residue in their active sites, and which play important roles in diverse physiological processes. including digestion (e.g. trypsin, chymotrypsin) and blood clotting (e.g. plasminogen activator, thrombin) Serine proteases also act as regulators of a variety of processes by proteolytic activation of precursor proteins.

The kallikreins are a sub-family of serine proteases originally defined as cleaving vasoactive peptides (kinins) from kininogen (Schachter M. (1980) *Pharmacol. Rev.* 31: 1 -1 7.). Currently the kallikreins comprise a large, multi-gene family in rodents, although only three members of this family are known in humans. These genes clustered on chromosome 19q13.2-q13.4 (Reigman PH, et al. (1992) *Genomics* 14:6-11) are hKLK1, hKLK2, and hKLK3 which encode the proteins hK1 (pancreatic/renal kallikrein), hK2 (glandular kallikrein), and hK3 (prostate specific antigen) respectively (Berg T, et al. (1992) *Agents Actions* 38 (Suppl 1):19-25).

The hK1 protein is secreted from pancreas, kidney, and salivary glands (Fukushima D,et al. (1985) *Biochemistry* 24:8037-8043), and is the only member of the family having true kallikrein activity. Its major function is the generation of kinins from kininogens and the regulation of blood pressure (Schachter , supra).

The hK2 protein has yet to be detected in human tissue or fluids, but its sequence has been inferred from a genomic clone (Schedlich LJ, et al. (1987) *DNA* 6:429-437) as well as cDNA clones isolated from prostate libraries (Schedlich LJ, et al., (1987)*DNA* 6:429-437). hK2 expression is specific for prostate and is regulated by androgens (Schedlich et al. supra). Determining the function for this protein and evaluating its usefulness as a marker for prostate cancer will have to await the identification and isolation of the protein.

The hK3 protein is PSA, the prostate specific antigen. It is produced predominantly in males by prostate epithelial cells and secreted into the seminal fluid where it serves to degrade the gel-like seminogelin protein and increase sperm motility (Lilja H. (1985) J. *Clin. Invest.* 76:1899-1903; Lilja H, et al. (1987) *J. Clin. Invest.* 80:281 -285). Although PSA is produced at higher levels in normal than in malignant prostate tissue, a defect in the malignant tissues ultimately results in the leakage of PSA into the bloodstream (McCormack RT, et al. (1995) *Urology* 45:729), forming the basis of the use of PSA as a marker for prostate cancer.

Serine proteases may accomplish matrix degradation during metastases by activating metalloproteases (Alexander and Werb. 1991. Extracellular Matrix Degradation. In Cell Biology of Extracellular Matrix. Ed by Hay, E.D. New York. Plenum Press. 1991:255). The principal serine proteases known implicated in matrix degradation mediate the plasminogen activation cascade. Included in this group are the urokinase plasminogen activator-receptor (uPA-uPAR) , leukocyte elastase, and tumor assocaited trypsin (Chen. 1992. *Curr. Opin. Cell Biol.* 4:802). Both uPA and tPA can activate serum protein plasminogen , yielding the broad-specificity protease plasmin by cleavage of one bond. Plasmin participates in fibrinolysis, tissue remodeling and tumor invasion (Chen, *supra*).

While proteases have been thought to promote tissue invasion and metastases, the development of metastatic potential appears to be more complicated. For example, overexpression of the protease inhibitors PAI-1 and PAI-2, which negatively regulate plasminogen activator, has also been found to be assocaited with certain types of cancers (Sumiyoshi et al. 1991 *Thromb. Res.*, 63:59; Reilly et al. 1990. *Biochem. Soc. Transact.* 18:354). Janicke et al. have hypothesized that increased PAI-1 secretion by tumor cells may enhance cell migration by upsetting the protease-antiprotease equilibrium near the cell surface of a tumor cell, perhaps via a mechanism involving urokinase plasminogen activator receptor clearance (Janicke et al. 1994. *Cancer Res.* 54:2527).

The identification of markers associated with the suppression of cancer, the development of cancer, and with the development of metastasis would be of great benefit.

Summary of the Invention

Disclosed herein is a novel member of the serine protease family, referred to as Protease M. A partial Protease M cDNA was originally identified by its differential expression in a primary ductal breast carcinoma and its reduced expression in a pleural metastasis from the same patient using the differential display method. Subsequently, a full-length cDNA of 1,526 nucleotides was isolated from a normal breast epithelial cell cDNA library and was sequenced. Expression studies indicate that expression of the Protease M gene is downregulated in metastatic breast cancer cell lines and is upregulated in primary breast cancer cell lines and ovarian cancer tissues and tumor cell lines.

In one aspect, this invention pertains to isolated nucleic acid molecules comprising a nucleotide sequence encoding a Protease M protein or a biologically active

portion thereof. In one embodiment, the invention features an isolated nucleic acid molecule comprising the nucleotide sequence shown in SEQ ID NO: 1.

In another embodiment the invention an isolated nucleic acid molecule of the present invention is at least 15 nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1.

In yet another embodiment, a nucleic acid molecule of the present invention comprises the coding region of the nucleotide sequence of SEQ ID NO: 1.

In still another embodiment the invention provides for isolated nucleic acid molecules which encode proteins containing amino acid sequences which are homologous to the sequence shown in SEQ ID NO:2. For example, in one embodiment, protein comprises an amino acid sequence at least 60 % homologous to the amino acid sequence of SEQ ID NO: 2. In another embodiment, the protein is at least about 70 %, preferably at least 80 % homologous, or more preferebly at least 90 % homologous to the amino acid sequence of SEQ ID NO: 2. In a preferred embodiment, an isolated nucleic acid molecule of the invention encodes the amino acid sequence of SEQ ID NO: 2.

In another embodiment, an isolated nucleic acid molecule encodes a Protease M fusion protein.

In yet another embodiment, an isolated nucleic acid molecule of the invention is antisense to the nucleic acid molecule of claim 1. In a preferred embodiment, an isolated nucleic acid is antisense to a coding region of the coding strand of the nucleotide sequence of SEQ ID NO: 1. In yet another embodiment, an isolated nucleic acid molecule of the invention is antisense to a noncoding region of the nucleotide sequence of SEQ ID NO: 1.

In one embodiment of the invention, an isolated nucleic acid molecule which encodes a Proteinase M polypeptide is isoloated using at least a portion of the nucleotide sequence of SEQ ID NO:1 as a probe or a primer.

Another aspect of the invention pertains to vectors, e.g., recombinant expression vectors, containing the nucleic acid molecules of the invention. Such vectors can encode a protein comprising the amino acid sequence of SEQ ID NO: 2. In one embodiment of the invention, a vector is provided which comprises the coding region of the nucleotide sequence of SEQ ID NO: 1. In one embodiment, such a host cell is used to produce Protease M protein by culturing the host cell in a suitable medium. If desired, Protease M protein can be then isolated from the medium or the host cell.

Still another aspect of the invention pertains to isolated Protease M protein. In one embodiment an isolated Protease M [illegible]

acid shown in SEQ ID No:1. In preferred embodiments, the Protease M protein is a mature polypeptide which comprises amino acids 17-244 of SEQ ID NO: 2 or amino acids 22-244. In other embodiments, the isolated Protease M protein comprises an amino acid sequence at least 60 % homologous to the amino acid sequence of SEQ ID NO: 2 and possesses a Protease M bioactivity *in vitro*. Preferably, the protein is at least 70 %, preferably at least 80 %, even more preferably at least 90% . In particularly preferred embodiments a Protease M protein of the present invention is at least about 95 % homologous to the amino acid sequence of SEQ ID NO: 2.

A Protease M protein of the invention can be incorporated into a pharmaceutical composition comprising the protein and a pharmaceutically acceptable carrier.

Moreover, the invention provides a fusion protein comprising a Protease M polypeptide operatively linked to a non-Protease M polypeptide.

The Protease M proteins of the invention, or fragments thereof, can be used to prepare anti-Protease M antibodies. The invention provides an antigenic peptide of Protease M comprising at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO: 2 and encompassing an epitope of Protease M such that an antibody raised against the peptide forms a specific immune complex with Protease M. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues. The invention further provides an antibody that specifically binds Protease M. In one embodiment, the antibody is monoclonal. In another embodiment, the antibody is coupled to a detectable label. In yet another embodiment, the antibody is incorporated into a pharmaceutical composition comprising the antibody and a pharmaceutically acceptable carrier.

Yet another aspect of the invention pertains to transgenic non-human animals in which a Protease M gene has been introduced or altered. In one embodiment, the genome of the nonhuman animal has been altered by introduction of a nucleic acid molecule of the invention encoding Protease M as a transgene. In another embodiment, an endogenous Protease M gene within the genome of the nonhuman animal has been altered, e.g., functionally disrupted, by homologous recombination.

Another aspect of the invention pertains to methods for detecting the presence or absence of Protease M in a biological sample. In a preferred embodiment, the method involves contacting a biological sample (e.g., a tissue sample) with an agent capable of detecting Protease M protein or nucleic acid such that the presence of Protease M is detected in the biological sample. The agent can be, for example, a labeled or labelable nucleic acid probe capable of hybridizing to Protease M mRNA or a

labeled or labelable antibody capable of binding to a Protease M protein. The invention further provides methods for detecting carcinomas or for staging a carcinoma based on detecting the presence, or absence, or amount of Protease M protein or nucleic acid in a test sample relative to a control sample. In one embodiment, the method involves contacting a cell or other sample from a subject with an agent capable of detecting Protease M protein or nucleic acid, determining the amount of Protease M protein or nucleic expressed in the sample, comparing the amount of Protease M protein or nucleic acid expressed in the sample to a control and forming a diagnosis and/or prognosis based on the amount of Protease M protein or nucleic acid expressed in the test sample as compared to the control sample. Preferably, the sample is mammary or ovarian tissue. For example, one such diagnostic method involves contacting the mRNA of a test cell with a nucleic acid probe containing a sequence antisense to (i.e. complementary to the sense strand of) a segment of the nucleic acid sequence shown in SEQ ID No:1. Kits for detecting Protease M in a biological sample are also within the scope of the invention.

The Protease M protein of the invention, and other agents related thereto, can be used therapeutically. For example the present invention can be used to modulate the Protease M bioactivity associated with a cell (e.g., in the cell, secreted by the cell or in the extracellular milieu surrounding the cell). Accordingly, in one embodiment, the invention provides a method for modulating the Protease M serine protease activity associated with a cell by contacting the cell with an agent that modulates Protease M serine protease activity. Such an agent can be, for example, a Protease M protein agonist or antagonist or a nucleic acid encoding a Protease M agonist or antagonist that has been introduced into the cell. In one embodiment, Protease M activity is stimulated in tumor cells, such as metastatic mammary tumor cells, in which endogenous Protease M expression is low or absent. Alternatively, in another embodiment, the invention provides a method for inhibiting the Protease M activity associated with a cell by contacting the cell with an agent that inhibits Protease M serine protease activity. Such an agent can be, for example, an antisense Protease M nucleic acid molecule or an anti-Protease M antibody, or Protease M antagonist, or inhibitor. The methods of the invention for modulating Protease M activity can be applied *in vitro* (e.g., to cells in culture) or *in vivo*, wherein an agent that modulates Protease M serine protease activity is administered to the subject. In a preferred embodiment, the invention provides a method for inhibiting development or progression of cancer in a cell comprising contacting a cell with an agent which modulates the amount of or activity of Protease M in or around the tumor cell.

Drug screening methods for identifying modulators of Protease M expression or Protease M serine protease activity are also encompassed by the invention. In one embodiment, the modulator stimulates Protease M expression or activity, i.e., is an agonist or potentiator. In another embodiment, the modulator inhibits Protease M expression or activity, i.e., is an antagonist or inhibitor.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Brief Description of the Drawings

Figure 1 shows the identification of Protease M (1G3) by Differential Display (DD) gel and northern blot (A.) DD gel: 21 PT and 21 MT-1 RNA was reverse transcribed with $T_{12}MG$ primer and PCR-amplified with $T_{12}MG$ and OPA1 primers in the presence of ^{35}SdATP, run on a 6% acrylamide sequencing gel, and exposed to x-ray film for 18 hours. The portion of the gel surrounding the differentially displayed 0.28kb band is shown. (B.) Northern Blot: 10mg of total cell RNA was northern blotted and probed with ^{32}P-labeled PCR-amplified 0.28kb band from the DD gel shown in (A).

Figure 2 shows Protease M cDNA. The cDNA sequence and putative protein coding sequence of the longest clone from the 76N library is shown. The postulated pre-pro N-terminal amino acids are underlined. The predicted cleavage sites of pre and pro amino acids after ala^{16} and lys^{21} respectively are indicated by arrows. The potential n-linked glycosylation site at amino acids 134-136 and asp^{191} at the bottom of the binding cleft are boxed. The residues of the catalytic triad (his^{62}, asp^{106}, and ser^{197}) are circled. The actual polyadenylation signal at nucleotide 1,490 and an alternative polyadenylation signal at nucleotide 1,095 are underlined.

Figure 3 shows an alignment of Protease M with closely related members of the serine protease family. The GCG pileup and pretty plot programs were used to align Protease M with closely related human serine proteases: They are from top to bottom: glandular kallikrein-hk2 (accession number SP|P06870|), PSA-hk3 (accession number SP|P07288|), pancreatic kallikrein-hk-1 (accession number SP|P20511|), and trypsinogen 1 (accession number SP|P07477|). Amino acids comprising the catalytic triad are marked with an asterisk. The 29 "invariant" amino acids (Dayhoff) are marked with a dot or an asterisk.

Figure 4 shows protease M mRNA expression in mammary and prostate cell lines (A.) 10mg of total mammary cell RNA was run on an agarose/ formaldehyde gel, blotted and hybridized to ^{32}P-labeled Protease M probe and exposed to x-ray film for 20 hours (B). 10mg of total prostate cell RNA was blotted and hybridized (as in A) and exposed to x-ray film for 20 hours.

Figure 5 shows Protease M mRNA expression in ovarian tissue.
10mg of total cell RNA isolated from ovarian tissue was blotted and hybridized to Protease M probe (as in Figure 4) and exposed to x-ray film for 5 days.

Figure 6 shows Protease M mRNA expression in human tissue.
A northern blot containing 2mg of polyA+ RNA from normal human tissue (Clontech) was hybridized to Protease M probes (as in Figure 4). The blot was exposed to x-ray film for 2 days.

Figure 7 shows the expression of Protease M protein in mammary cell lines and insect cells infected with recombinant Protease M . 50mg of total cell lysate from mammary cell lines, uninfected insect cells (SF9) or insect cells infected with 4.5ml recombinant Protease M baculovirus (SF9/1G3(1)) or 22.5ml recombinant baculovirus (SF9/Protease M(2)) was run on a 12% polyacryamide/SDS gel, transferred to a PDVF membrane, and reacted with Protease M polyclonal anti-peptide antibody as the primary antibody and horseradish peroxidase conjugated anti rabbit IgG secondary antibody. Bands were detected with ECL detection system.

Detailed Description of the Invention

Protease M was isolated by differential display (Liang L and Pardee AB. (1992) *Science* 257:967-970; Liang L, et al.. (1993) *Nucleic Acids Res.* 21:32673275; Sager R, et al. (1993) *FASEB J.* 7: 964-970). Protease M is a novel member of the serine protease family which is most homologous to trypsin and members of the kallikrein family. Protease M is downregulated in metastatic breast cancer lines, but strongly expressed at the mRNA level in some primary breast cancer cell lines and in ovarian cancer tissues and tumor cell lines.

Protease M was originally identified as being differentially expressed in a primary ductal breast carcinoma (21PT) as compared to a pleural metastasis (21MT-1) derived from the same patient. A full-length cDNA was subsequently isolated using the

partial cDNA as a hybridization probe to screen a cDNA library prepared from a normal breast epithelial cell (76N).

The nucleotide sequence of the isolated human Protease M cDNA, and the predicted amino acid sequence of the human Protease M protein, are shown in SEQ ID NOs: 1 and 2, respectively. The full length cDNA clone isolated is 1526 nucleotides in length and comprises 2456 base pairs of 5' nontranslated sequence, 732 base pairs of coding sequence, and 549 base pairs of 3' nontranslated sequence. The predicted Protease M protein is 244 amino acids. The NH_2 terminus comprises 13 consecutive hydrophobic amino acids (leu^4-ala^{16}), which is a predicted signal sequence. The residues glu^{17}-glu^{18}- glu^{19}-asn^{20}-lys^{21} resemble a pro-polypeptide with a potential trypsin cleavage site after lys^{21}.

Comparison of Protease M with other known proteins showed that Glandular kallikrein 2 (Schedlich LJ, et al. (1987) *DNA* 6:429-437; Riegman PH, et al. (1991) *Mol. Cell Endocrinol.* 76:181-190) has 44% exact matches and 48% match with conservative changes. Trypsin I (Emi M, et al. (1986) *Gene* 41:305-310) has 43% exact matches and 49% match with conservative changes. Both glandular kallikrein 1 (Fukushima D, et al. (1985) *Biochemistry* 24:8037-8043., Baker A, Shine J. (1985)*DNA* 4:445 -450; Takahashi S, Irie A, Miyake Y. (1988) *Biochem.* 104:22-29; Lu HS, et al. (1989) *Int. J. Peptide Protein Res.* 33:237 -249; Angermann A, et al. (1989). *Biochem. J. 262:787*793) and prostate specific antigen (Watt, et al. (1986) *Proc. Natl. Acad. Sci.* USA 83:3166-3170; Lundwall A, Lilja H. (1987) *FEBS Letters* 214:317-322; Schaller J, et al. (1987) *Eur. J.Biochem.* 170:111-120; Riegman PHJ, Klaassen P, et al. (1988) *Biochem. and Biophys. Res. Comm.* 155: 1 8 1 - 188; Henttu P andVihko P. (1989).*Biochem. and Biophys. Res. Comm* 60:903-910) have 39% exact matches and 44% match with conservative changes.

Structural features important for serine protease activity such as the catalytic triad ($his^{62}asn^{106}ser^{197}$), cysteine bridges ($Cys^{28}$-$Cys^{157}$; Cys^{47}-Cys^{63}; Cys^{138}-Cys^{203}; Cys^{168}-Cys^{182}; and Cys^{193}-Cys^{218}), and residues lining the binding cleft are almost perfectly conserved between Protease M and other members of the kallikrein family. The Asp residue at position 191 predicts that Protease M has a trypsin-like cleavage pattern. Unlike the members of the kallikrein family, Protease M and trypsin lack the kallikrein loop at amino acid residues 109-119, which is important for kallikrein specificity.

Moreover, Protease M mRNA has a distinct expression pattern that distinguishes it from other serine proteases. A 1.7-1.8 kb message was found to be normal brain, kidney, and pancreas tissue, but not in heart, placenta, lung, liver, or skeletal muscle. The message detected in the pancreas was only about 1.2 kb.

Expression studies further indicate that expression of the Protease M gene is downregulated in metastatic breast cancer cell lines and is upregulated in primary breast cancer cell lines and ovarian cancer tissues and tumor cell lines.

The Protease M gene was localized by FISH analysis to chromosome 19q13.4. The three kallikrein genes also map to chromosome 19q13.2-q13.4, while trypsinogen 1 maps to chromosome 7. These mapping data suggest that Protease M is probably more closely related on an evolutionary basis to the kallikreins than to trypsin.

The size of the detected Protease M protein is approximately 36 kD rather than the predicted size of 27 kD. This size discrepancy could be accounted for by glycosylation at asn^{134}. The expression of Protease M is regulated both at the transcriptional and translational level.

Accordingly, certain aspects of the present invention relate to nucleic acids encoding Protease M proteins, the Protease M proteins themselves, antibodies immunoreactive with Protease M proteins, and preparations of such compositions. Moreover, the present invention provides diagnostic/prognostic assays and therapeutic reagents for detecting and treating disorders involving, for example, aberrant expression of Protease M or Protease M homologs. In addition, drug discovery assays are provided for identifying agents which can modulate the biological function of Protease M proteins, such as by altering the binding of Protease M molecules to proteins, including substrates. Such agents can be useful therapeutically to alter the growth and/or differentiation of a cell. Other aspects of the invention are described below or will be apparent to those skilled in the art in light of the present disclosure.

Various aspects of the invention are described in further detail in the following subsections:

I. Definitions

In general, polypeptides referred to herein as having an activity of a Protease M protein (e.g., are "bioactive") are defined as polypeptides which include an amino acid sequence corresponding (e.g., identical or homologous) to all or a portion of the amino acid sequences of a Protease M protein shown in SEQ ID No:2 and which mimic or antagonize all or a portion of the biological/biochemical activities of a naturally occurring Protease M protein. Examples of such biological activity include serine protease activity and/or the ability to compete with a bioactivity of a naturally occurring Protease M. The ability of portions of Protease M to exhibit serine protease activity can be determined in standard *in vitro* serine protease assays, for example as described in detail in the appended examples. In other embodiments, a Protease M

molecule of the present invention is capable of modulating the proliferation or metastasis of a cell, either *in vitro or in vivo*.

Other biological activities of the subject Protease M proteins are described herein or will be reasonably apparent to the skilled artisan. According to the present invention, a polypeptide has biological activity if it is a specific agonist or antagonist of a naturally-occurring form of a Protease M protein.

"Cells," "host cells" or "recombinant host cells" are terms used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A "chimeric protein" or "fusion protein" is a fusion of a first amino acid sequence encoding one of the subject Protease M polypeptides with a second amino acid sequence defining a domain (e.g. polypeptide portion) foreign to and not substantially homologous with any domain of one of the Protease M proteins. A chimeric protein may present a foreign domain which is found (albeit in a different protein) in an organism which also expresses the first protein, or it may be an "interspecies", "intergenic", etc. fusion of protein structures expressed by different kinds of organisms. In general, a fusion protein can be represented by the general formula X-Protease M-Y, wherein Protease M represents a portion of the protein which is derived from a Protease M protein, and X and Y are, independently, absent or represent amino acid sequences which are not related to a Protease M sequence in an organism.

As is well known, genes for a particular polypeptide may exist in single or multiple copies within the genome of an individual. Such duplicate genes may be identical or may have certain modifications, including nucleotide substitutions, additions or deletions, which all still code for polypeptides having substantially the same activity. The term "DNA sequence encoding a Protease M polypeptide" may thus refer to one or more genes within a particular individual. Moreover, certain differences in nucleotide sequences may exist between individuals of the same species, which are called alleles. Such allelic differences may or may not result in differences in amino acid sequence of the encoded polypeptide yet still encode a protein with the same biological activity.

As used herein, the term "gene" or "recombinant gene" refers to a nucleic acid comprising an open reading frame encoding a Protease M polypeptide of the present invention, including both exon and (optionally) intron sequences. A "recombinant gene" refers to nucleic acid encoding a Protease M polypeptide and comprising Protease M-encoding exon sequences, though it may optionally include

molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA). Nucleic acids may be double stranded or single stranded and the term is meant to include a nucleic acid which is complementary (i.e., can specifically hybridize to) a nucleic acid of the present invention (e.g., an antisense molecule). The term "nucleic acid" as used herein is intended to include fragments as equivalents. The term equivalent is understood to include nucleotide sequences encoding functionally equivalent Protease M polypeptides or functionally equivalent peptides having a bioactivity of a Protease M protein such as described herein. Equivalent nucleotide sequences will include sequences that differ by one or more nucleotide substitutions, additions or deletions, such as allelic variants; and will, therefore, include sequences that differ from the nucleotide sequence of the Protease M cDNA sequences shown in SEQ ID No:1 due to the degeneracy of the genetic code. Equivalents will also include nucleotide sequences that hybridize under stringent conditions (i.e., equivalent to about 20-27°C below the melting temperature (T_m) of the DNA duplex formed in about 1M salt) to the nucleotide sequence represented in SEQ ID No:1. In one embodiment, equivalents will further include nucleic acid sequences derived from and evolutionarily related to, a nucleotide sequence shown in SEQ ID No:1.

As used herein, the term "specifically hybridizes" refers to the ability of the probe/primer of the invention to hybridize to at least 15 consecutive nucleotides of a Protease M gene, such as a Protease M sequence designated in SEQ ID No: 1, or a sequence complementary thereto, or naturally occurring mutants thereof, such that it has less than 15%, preferably less than 10%, and more preferably less than 5% background hybridization to a cellular nucleic acid (e.g., mRNA or genomic DNA) encoding a protein other than a Protease M protein, as defined herein.

As used herein, the term "tissue-specific promoter" means a DNA sequence that serves as a promoter, i.e., regulates expression of a selected DNA sequence operably linked to the promoter, and which effects expression of the selected DNA sequence in specific cells of a tissue, such as cells of hepatic, pancreatic, neuronal or hematopoietic origin. The term also covers so-called "leaky" promoters, which regulate expression of a selected DNA primarily in one tissue, but can cause at least low level expression in other tissues as well.

As used herein, the term "transfection" means the introduction of a nucleic acid, e.g., an expression vector, into a recipient cell by nucleic acid-mediated gene transfer. "Transformation", as used herein, refers to a process in which a cell's genotype is changed as a result of the cellular uptake of exogenous DNA or RNA, and, for example, the transformed cell expresses a recombinant form of a Protease M

polypeptide or, where anti-sense expression occurs from the transferred gene, the expression of a naturally-occurring form of the Protease M protein is disrupted.

As used herein, a "transgenic animal" is any animal, preferably a non-human mammal, bird or an amphibian, in which one or more of the cells of the animal contain heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA. In the typical transgenic animals described herein, the transgene causes cells to express a recombinant form of a Protease M protein. However, transgenic animals in which the recombinant Protease M gene is silent are also contemplated, as for example, the FLP or CRE recombinase dependent constructs described below. Moreover, "transgenic animal" also includes those recombinant animals in which gene disruption of one or more Protease M genes is caused by human intervention, including both recombination and antisense techniques.

"Transcriptional regulatory sequence" is a generic term used throughout the specification to refer to DNA sequences, such as initiation signals, enhancers, and promoters, which induce or control transcription of protein coding sequences with which they are operably linked. In preferred embodiments, transcription of a recombinant Protease M gene is under the control of a promoter sequence (or other transcriptional regulatory sequence) which controls the expression of the recombinant gene in a cell-type in which expression is intended. It will also be understood that the recombinant gene can be under the control of transcriptional regulatory sequences which are the same or which are different from those sequences which control transcription of naturally-occurring forms of Protease M genes.

As used herein, the term "transgene" means a nucleic acid sequence (encoding a Protease M polypeptide, or an antisense transcript thereto), which is partly or entirely heterologous, i.e., foreign, to the transgenic animal or cell into which it is introduced, or, is homologous to an endogenous gene of the transgenic animal or cell into which it is introduced, but which is designed to be inserted, or is inserted, into the animal's genome in such a way as to alter the genome of the cell into which it is inserted (e.g., it is inserted at a location which differs from that of the natural gene or its insertion results in a knockout). A transgene can include one or more transcriptional regulatory

sequences and any other nucleic acid, such as introns, that may be necessary for optimal expression of a selected nucleic acid.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of preferred vector is an episome, i.e., a nucleic acid capable of extra-chromosomal replication. Preferred vectors are those capable of autonomous replication and/expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of "plasmids" which refer generally to circular double stranded DNA loops which, in their vector form are not bound to the chromosome. In the present specification, "plasmid" and "vector" are used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors which serve equivalent functions and which become known in the art subsequently hereto.

II. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode Protease M or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes to identify Protease M-encoding nucleic acid.

A Protease M nucleic acid or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, a human Protease M cDNA can be isolated from a cell line, (e.g., a normal mammary epithelial cell line) or from a cDNA library, using all or portion of SEQ ID NO: 1 as a hybridization probe and standard hybridization techniques (e.g., as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor, NY, 1989). Moreover, a nucleic acid molecule encompassing all or a portion of SEQ ID NO: 1 can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon the sequence of SEQ ID NO: 1. For example, mRNA can be isolated from normal mammary epithelial cells (e.g., by the guanidinium-thiocyanate extraction procedure of Chirgwin et al. (1979) *Biochemistry* 18: 5294-5299) and cDNA can be prepared using reverse transcriptase (e.g., Moloney MLV reverse transcriptase, available from Gibco/BRL, Bethesda, MD; or AMV reverse transcriptase, available from Seikagaku America, Inc., St. Petersburg, FL). Synthetic oligonucleotide primers for PCR amplification can be designed based upon the nucleotide sequence shown in SEQ

ID NO: 1. For example, primers suitable for amplification of a Protease M nucleic acid are provided in the appended Examples. A nucleic acid of the invention can be amplified using cDNA or, alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to Protease M nucleotide sequence can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In one embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO: 1 or a fragment thereof. The sequence of SEQ ID NO: 1 corresponds to the human Protease M cDNA. This cDNA comprises sequences encoding the Protease M protein (i.e., "the coding region", from nucleotides 246 to 977), as well as 5' untranslated sequences (nucleotides 1 to 245) and 3' untranslated sequences (nucleotides 978 to 1526). Alternatively, the nucleic acid molecule may comprise only the coding region of SEQ ID NO: 1 (e.g., nucleotides 246 to 977), for example a fragment encoding a biologically active portion of Protease M.

In another embodiment, the Protease M nucleic acid of the present invention encodes the polypeptide shown in SEQ ID No:2. In another embodiment, a Protease M nucleic acid encodes a biologically active portion of Protease M. In yet another embodiment a Protease M nucleic acid encodes a mature form of Protease M in which a hydrophobic, amino-terminal signal sequence (encompassing approximately amino acids 1-16) is absent. In a further embodiment, a mature form of Protease M preferably comprises about amino acid residues 22 to 244 (i.e., Protease M which has been cleaved at a trypsin site). Although, in preferred embodiments, a nucleic acid of the present invention encodes a protein in which amino acid residue 22 is the N-terminal residue of the mature protein, more than one native isoform differing in the length of the N-terminal sequence may exist for Protease M. Consequently, the skilled artisan will appreciate that some flexibility exists in the N-terminus of the mature form of Protease M lacking a signal sequence. Additional nucleic acid fragments encoding biologically active portions of Protease M can be prepared by isolating a portion of SEQ ID NO: 1, expressing the encoded portion of Protease M protein or peptide (e.g., by recombinant expression *in vitro*) and assessing the bioactivity of the encoded portion of Protease M protein or peptide.

In another embodiment, an isolated nucleic acid molecule of the invention is at least 15 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1. In other embodiment, the nucleic acid is at least 30, 50, 100, 250 or 500 nucleotides in

length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60 % homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that at least sequences at least 65 %, more preferably at least 70 %, and even more preferably at least 75 % homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO: 1 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein). In one embodiment, the nucleic acid encodes a natural human Protease M. In another embodiment, the nucleic acid molecule encodes a murine homologue of human Protease M.

In one embodiment a Protease M nucleic acid of the present invention comprises the sequence shown in SEQ ID NO: 1 or a fragment thereof. It will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of Protease M may exist within a population (e.g., the human population). Such genetic polymorphism in the Protease M gene may exist among individuals within a population due to natural allelic variation. Such natural allelic variations can typically result in 1-5 % variance in the nucleotide sequence of a gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in Protease M that are the result of natural allelic variation and that do not alter the functional activity of Protease M are within the scope of the invention. Moreover, nucleic acid molecules encoding Protease M proteins from other species, and thus which have a nucleotide sequence which differs from the human sequence of SEQ ID NO: 1, are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and nonhuman homologues of the human Protease M cDNA of the invention can be isolated based on their homology to the human Protease M nucleic acid disclosed herein using the human cDNA, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

In addition to naturally-occurring allelic variants of the Protease M

changes may be introduced by mutation into the nucleotide sequence of SEQ ID NO: 1, thereby leading to changes in the amino acid sequence of the encoded Protease M protein, without altering the functional ability of the Protease M protein, as described in more detail below. Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding Protease M proteins that contain changes in amino acid residues that are not essential for Protease M activity , e.g., residues that are not conserved or only semi-conserved among members of the chymotrypsin family of serine proteases. Such Protease M proteins differ in amino acid sequence from SEQ ID NO: 2 yet retain Protease M bioactivity.

The invention further encompasses nucleic acid molecules that differ from SEQ ID NO:1 (and portions thereof) due to degeneracy of the genetic code. In another embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least 60 % homologous to the amino acid sequence of SEQ ID NO: 2 and exhibits serine protease activity *in vitro*. Preferably, the protein encoded by the nucleic acid molecule is at least 70 % homologous to SEQ ID NO: 2, more preferably at least 80 % homologous to SEQ ID NO: 2, even more preferably at least 90 % homologous to SEQ ID NO: 2. In a particularly preferred embodiment a Protease M nucleic acid of the present invention is at least about 95 % homologous to SEQ ID NO: 2.

To determine the percent homology of two amino acid sequences (e.g., SEQ ID NO: 2 and a mutant form thereof), the sequences are aligned for optimal comparison purposes (e.g., gaps may be introduced in the sequence of one protein for optimal alignment with the other protein). The amino acid residues at corresponding amino acid positions are then compared. When a position in one sequence (e.g., SEQ ID NO: 2) is occupied by the same amino acid residue as the corresponding position in the other sequence (e.g., a mutant form of Protease M), then the molecules are homologous at that position (i.e., as used herein amino acid "homology" is equivalent to amino acid "identity"). The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % homology = # of identical positions/total # of positions x 100).

An isolated nucleic acid molecule encoding a Protease M protein homologous to the protein of SEQ ID NO: 2 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO: 1 such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein, as detailed below.

In addition to the nucleic acid molecules encoding Protease M proteins described above, another aspect of the invention pertains to isolated nucleic acid

molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid.

The antisense nucleic acid can be complementary to an entire Protease M coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding Protease M. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (e.g., the entire coding region of SEQ ID NO: 1 comprises nucleotides 246 to 977). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding Protease M. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding Protease M disclosed herein (e.g., SEQ ID NO: 1), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. Preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of Protease M mRNA. For example, the antisense oligonucleotide may be complementary to the region surrounding the translation start site of Protease M mRNA. An antisense oligonucleotide can be, for example, about 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

In another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they

nucleic acid can be designed based upon the nucleotide sequence of a Protease M cDNA disclosed herein (i.e., SEQ ID NO: 1). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the base sequence of the active site is complementary to the base sequence to be cleaved in a Protease M-encoding mRNA. See for example Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742. Alternatively, Protease M mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See for example Bartel, D. and Szostak, J.W. (1993) *Science* 261: 1411-1418.

III. Recombinant Expression Vectors and Host Cells

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., Protease M proteins, mutant forms of Protease M, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of Protease M in prokaryotic or eukaryotic cells. For example, Protease M can be expressed in bacterial cells such as *E. coli* or insect cells (using baculovirus expression vectors) as described in detain in the appended Examples. Other possible host cells include yeast cells or mammalian cells. Suitable host cells are discussed

further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector may be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. In a preferred embodiment, exemplified herein, the coding sequence of the mature form of Protease M (i.e., encompassing amino acids 22-244) is cloned into a pGEX-2t expression vector to create a vector encoding a fusion protein which was solubilized from bacteria and purified on glutathionine agarose beads by standard methods (Smith DB, and Johnson. 1988. *Gene* 67:31).

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., (1988) *Gene* 69:301-315) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave

Enzymology 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada et al., (1992) *Nuc. Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the Protease M expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1 (Baldari. et al., (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz et al., (1987) *Gene* 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA).

Alternatively, Protease M can be expressed in insect cells using baculovirus expression vectors as described herein. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al., (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow, V.A., and Summers, M.D., (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B., (1987) *Nature* 329:840) and pMT2PC (Kaufman et al. (1987), *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al. (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters

(Kessel and Gruss (1990) *Science* 249:374-379) and the a-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to Protease M mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. et al., Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to recombinant host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell may be any prokaryotic or eukaryotic cell. For example, Protease M protein may be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or

Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory press (1989)), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker may be introduced into a host cell on the same vector as that encoding protease M or may be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) Protease M protein. Accordingly, the invention further provides methods for producing Protease M protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding Protease M has been introduced) in a suitable medium until Protease M is produced. In another embodiment, the method further comprises isolating Protease M from the medium or the host cell.

IV. Protease M Proteins

Another aspect of the invention pertains to isolated Protease M proteins, and biologically active portions thereof, as well as peptide fragments suitable as immunogens to raise anti-Protease M antibodies. The invention provides an isolated preparation of Protease M, or a biologically active portion thereof. An "isolated" protein is substantially free of cellular material or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. In a preferred embodiment, the Protease M protein has an amino acid sequence shown in SEQ ID NO: 2. In other embodiments, the Protease M protein is substantially homologous to SEQ ID NO: 2 and retains the functional activity of the protein of SEQ ID NO: 2 yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the Protease M protein is a protein which comprises an amino acid sequence at least 60 % homologous to the amino acid sequence of SEQ ID NO: 2 and possesses a Protease M bioactivity *in vitro*. Preferably, the protein is at least 70 % homologous to SEQ ID NO: 2, more preferably at least 80 % homologous to SEQ ID

NO: 2, even more preferably at least 90 % homologous to SEQ ID NO: 2. In a particularly preferred embodiment, a Protease M polypeptide is at least about 95 % homologous to SEQ ID NO: 2.

An isolated Protease M protein may comprise the entire amino acid sequence of SEQ ID NO: 2 (i.e., amino acids 1-244) or a biologically active portion thereof. For example, a biologically active portion of Protease M can comprise a mature form of Protease M in which a hydrophobic, amino-terminal signal sequence is absent, or which has been cleaved at a trypsin site. In one embodiment, such a mature form of Protease M comprises about amino acids 17-244 of SEQ ID NO: 2 and in another embodiment comprises about amino acids 22-244. The term "about" amino acids 17-244 or 22-244 is intended to indicate that there is some flexibility in the amino-terminal residue, as discussed further in subsection I above. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for serine protease activity as described in detail above.

Protease M proteins are preferably produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the protein is cloned into an expression vector (as described above), the expression vector is introduced into a host cell (as described above) and the Protease M protein is expressed in the host cell. The Protease M protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Alternative to recombinant expression, a Protease M protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques. Moreover, native Protease M protein can be isolated from cells (e.g., cultured human mammary epithelial cells), for example using an anti-Protease M antibody (discussed further below).

In yet another embodiment of the present invention a Protease M protein is encoded by a nucleic acid of SEQ ID No:1. In another embodiment a Protease M protein is encoded by a nucleic acid at least about 60%, preferably about 70%, or more preferably about 80% homologous to the nucleic acid of SEQ ID No:1. In a particularly preferred embodiment a Protease M protein is encoded by a nucleic acid at least about 90% and preferably about 95% homologous to the nucleic acid of SEQ ID No:1. In still another embodiment of the present invention a Protease M polypeptide is encoded by a nucleic acid which hybridizes to the nucleic acid of SEQ ID No:1 under stringent conditions.

The invention also provides Protease M fusion proteins. As used herein, a Protease M "fusion protein" comprises a Protease M polypeptide operatively linked to a non-Protease M polypeptide. An "Protease M polypeptide" refers to a polypeptide

having an amino acid sequence corresponding to Protease M, whereas a "non-Protease M polypeptide" refers to a polypeptide having an amino acid sequence corresponding to another protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the Protease M polypeptide and the non-Protease M polypeptide are fused in-frame to each other. The non-Protease M polypeptide may be fused to the N-terminus or C-terminus of the Protease M polypeptide. For example, in one embodiment the fusion protein is a GST-Protease M fusion protein in which the Protease M sequences are fused to the C-terminus of the GST sequences (see Example 3). Such fusion proteins can facilitate the purification of recombinant Protease M. In another embodiment, the fusion protein is a Protease M protein containing a heterologous signal sequence at its N-terminus. For example, the native Protease M signal sequence (i.e., about amino acids 1-16) can be removed and replaced with a signal sequence from another protein. In certain host cells (e.g., mammalian host cells), expression and/or secretion of Protease M may be increased through use of a heterologous signal sequence.

Preferably, a Protease M fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel et al. John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A Protease M-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the Protease M protein.

An isolated Protease M protein, or fragment thereof, can be used as an immunogen to generate antibodies that bind Protease M using standard techniques for polyclonal and monoclonal antibody preparation. In particularly preferred embodiments, the Protease M immunogen comprises an epitope unique to Protease M. The full-length Protease M protein can be used or, alternatively, the invention provides antigenic peptide fragments of Protease M for use as immunogens. The antigenic peptide of Protease M comprises at least 8 amino acid residues of the amino acid

sequence shown in SEQ ID NO: 2 and encompasses an epitope of Protease M such that an antibody raised against the peptide forms a specific immune complex with Protease M. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of Protease M that are located on the surface of the protein, e.g., hydrophilic regions. Exemplary immunogens are described in more detail in the appended examples.

A Protease M immunogen typically is used to prepare antibodies by immunizing a suitable subject, (e.g., rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed Protease M protein or a chemically synthesized Protease M peptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic Protease M preparation induces a polyclonal anti-Protease M antibody response.

Modification of the structure of the subject Protease M polypeptides can be for such purposes as enhancing therapeutic or prophylactic efficacy, stability (e.g., *ex vivo* shelf life and resistance to proteolytic degradation *in vivo*), or post-translational modifications. Such modified peptides, when designed to retain at least one activity of the naturally-occurring form of the protein, or to produce specific antagonists thereof, are considered functional equivalents of the Protease M polypeptides described in more detail herein. Such modified peptides can be produced, for instance, by amino acid substitution, deletion, or addition.

For example, it is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid (i.e. isosteric and/or isoelectric mutations) will not have a major effect on the biological activity of the resulting molecule. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids are can be divided into four families: (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine, histidine; (3) nonpolar = alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar = glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. In similar fashion, the amino acid repertoire can be grouped as (1) acidic = aspartate, glutamate;

isoleucine, serine, threonine, with serine and threonine optionally be grouped separately as aliphatic-hydroxyl; (4) aromatic = phenylalanine, tyrosine, tryptophan; (5) amide = asparagine, glutamine; and (6) sulfur -containing = cysteine and methionine. (see, for example, Biochemistry, 2nd ed., Ed. by L. Stryer, WH Freeman and Co.: 1981). Whether a change in the amino acid sequence of a peptide results in a functional Protease M homolog (e.g. functional in the sense that the resulting polypeptide mimics or antagonizes the wild-type form) can be readily determined by assessing the ability of the variant peptide to produce a response in cells in a fashion similar to the wild-type protein, or competitively inhibit such a response. Polypeptides in which more than one replacement has taken place can readily be tested in the same manner.

Amino acid residues of Protease M that are strongly conserved among members of the chymotrypsin family of serine proteases (e.g., amino acid residues involved in substrate catalysis) are predicted to be essential to the bioactivity of Protease M and thus are not likely to be amenable to alteration. For example, the catalytic residues of a serine protease are Ser^{195}, His^{57} and Asp^{102} (chymotrypsin numbering system). These three residues form a hydrogen bonding system often referred to as the catalytic triad, or the charge relay system (Powers and Harper, supra). The catalytic triad of serine proteases is conserved in Protease M (i.e. $histidine^{62}$, $aspartate^{106}$, and $serine^{197}$). The aspartate at position 191 predicts that this protein will produce trypsin-like cleavage, and likewise, this residue may not be amenable to alteration.

Protease M contains twelve cysteine residues. Ten of these are conserved in the two kallikreins, PSA and human trypsin and would be expected to form the following disulfide bridges: (Cys^{28}-Cys^{157}). (Cys^{47}-Cys^{63}), (Cys^{138}-Cys^{203}), (Cys^{168}-Cys^{182}), and (Cys^{193}-Cys^{218}). The other two cysteines (Cys^{131} and Cys^{231}) are not found in the kallikreins, PSA and human trypsin, but are found in similar positions in bovine trypsin and would be expected to form a disulfide bond.

Twenty seven of the twenty nine 'invariant' amino acids surrounding the active site of serine proteases (Dayhoff MO. (1978) *Natl. Biomed. Res. Found.*, Washington DC, 5:Suppl. 3, pp. 79-81.) are conserved in Protease M. One of the two nonconserved amino acids $ileu^{155}$ in Protease M in place of leu is a conservative change. The other nonconserved amino acid, his^{161} in Protease M instead of pro, is also found in glandular kallikrein and PSA. The kallikreins and PSA have 11 amino acids residues 109-119 which are not found in Protease M or trypsin. The function of these amino acids is not clear, but they would be expected to form the so called kallikrein loop which would determine substrate specificity (Ashley PL, MacDonald RJ. (1985) *Biochemistry* 24:45 2-45 20.). Given their conservation among the serine proteases, they are likely to be important in the bioactivity of protease M. Other

important amino acid residues are predicted to be those which are conserved in 4 of the 5 serine proteases shown in Figure 3.

This invention further contemplates a method for generating sets of combinatorial mutants of the subject Protease M proteins as well as truncation mutants, and is especially useful for identifying potential variant sequences (e.g. homologs) that have a Protease M activity. The purpose of screening such combinatorial libraries is to generate, for example, novel Protease M homologs which can act as either agonists or antagonists, or alternatively, possess novel activities all together. To illustrate, Protease M homologs can be engineered by the present method to provide selective, constitutive activation of enzymatic activity. Thus, combinatorially-derived homologs can be generated to have an increased potency relative to a naturally occurring form of the protein.

Likewise, Protease M homologs can be generated by the present combinatorial approach to selectively inhibit (antagonize) a Protease M activity. For instance, mutagenesis can provide Protease M homologs which are able to prevent serine protease activity, e.g. the homologs can be dominant negative mutants. In a preferred embodiment, a dominant negative mutant of a Protease M protein is mutated at one or more residues of its catalytic site and/or specificity subsites.

In one aspect of this method, the amino acid sequences for a population of Protease M homologs or other related proteins are aligned, preferably to promote the highest homology possible. Such a population of variants can include, for example, Protease M homologs from one or more species. Amino acids which appear at each position of the aligned sequences are selected to create a degenerate set of combinatorial sequences. In a preferred embodiment, the variegated library of Protease M variants is generated by combinatorial mutagenesis at the nucleic acid level, and is encoded by a variegated gene library. For instance, a mixture of synthetic oligonucleotides can be enzymatically ligated into gene sequences such that the degenerate set of potential Protease M sequences are expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g. for phage display) containing the set of Protease M sequences therein.

There are many ways by which such libraries of potential Protease M homologs can be generated from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be carried out in an automatic DNA synthesizer, and the synthetic genes then ligated into an appropriate expression vector. The purpose of a degenerate set of genes is to provide, in one mixture, all of the sequences encoding the desired set of potential Protease M sequences. The synthesis of degenerate oligonucleotides is well known in the art (see for example, Narang, S.A.

(1983) Tetrahedron 39:3; Itakura et al. (1981) Recombinant DNA, Proc 3rd Cleveland Sympos. Macromolecules, ed. AG Walton, Amsterdam: Elsevier pp273-289; Itakura et al. (1984) Annu. Rev. Biochem. 53:323; Itakura et al. (1984) Science 198:1056; Ike et al. (1983) Nucleic Acid Res. 11:477. Such techniques have been employed in the directed evolution of other proteins (see, for example, Scott et al. (1990) Science 249:386-390; Roberts et al. (1992) PNAS 89:2429-2433; Devlin et al. (1990) Science 249: 404-406; Cwirla et al. (1990) PNAS 87: 6378-6382; as well as U.S. Patents Nos. 5,223,409, 5,198,346, and 5,096,815).

Likewise, a library of coding sequence fragments can be provided for a Protease M clone in order to generate a variegated population of Protease M fragments for screening and subsequent selection of bioactive fragments. A variety of techniques are known in the art for generating such libraries, including chemical synthesis. In one embodiment, a library of coding sequence fragments can be generated by (i) treating a double stranded PCR fragment of a Protease M coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule; (ii) denaturing the double stranded DNA; (iii) renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products; (iv) removing single stranded portions from reformed duplexes by treatment with S1 nuclease; and (v) ligating the resulting fragment library into an expression vector. By this exemplary method, an expression library can be derived which codes for N-terminal, C-terminal and internal fragments of various sizes.

A wide range of techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a certain property. Such techniques will be generally adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of Protease M homologs. The most widely used techniques for screening large gene libraries typically comprises cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates relatively easy isolation of the vector encoding the gene whose product was detected.

In an exemplary embodiment, the library of Protease M variants is expressed as a fusion protein on the surface of a viral particle. For instance, in the filamentous phage system, foreign peptide sequences can be expressed on the surface of infectious phage, thereby conferring two significant benefits. First, since these phage can be applied to affinity matrices at very high concentrations, a large number of phage can be screened at one time. Second, since each infectious phage displays the

combinatorial gene product on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be amplified by another round of infection. The group of almost identical *E. coli* filamentous phages M13, fd., and f1 are most often used in phage display libraries, as either of the phage gIII or gVIII coat proteins can be used to generate fusion proteins without disrupting the ultimate packaging of the viral particle (Ladner et al. PCT publication WO 90/02909; Garrard et al., PCT publication WO 92/09690; Marks et al. (1992) *J. Biol. Chem.* 267:16007-16010; Griffiths et al. (1993) *EMBO J* 12:725-734; Clackson et al. (1991) *Nature* 352:624-628; and Barbas et al. (1992) *PNAS* 89:4457-4461).

For example, the recombinant phage antibody system (RPAS, Pharmacia Catalog number 27-9400-01) can be easily modified for use in expressing and screening Protease M combinatorial libraries by panning on glutathione immobilized substrate/GST fusion proteins to enrich for Protease M homologs which retain an ability to bind a substrate or regulatory protein. Each of these Protease M homologs can subsequently be screened for further biological activities in order to differentiate agonists and antagonists. For example, homologs isolated from the combinatorial library can be tested for their enzymatic activity directly, or for their effect on cellular proliferation relative to the wild-type form of the protein.

The invention also provides for reduction of the Protease M proteins to generate mimetics, e.g. peptide or non-peptide agents, which are able to disrupt a biological activity of a Protease M polypeptide of the present invention, e.g. as catalytic inhibitor or an inhibitor of protein-protein interactions. Thus, such mutagenic techniques as described above are also useful to map the determinants of the Protease M proteins which participate in protein-protein interactions. To illustrate, the critical residues of a subject Protease M polypeptide which are involved proteolytic cleavage can be used to generate Protease M-derived peptidomimetics which competitively inhibit binding of the authentic Protease M protein with that moiety. By employing, for example, scanning mutagenesis to map the amino acid residues of a protein which is involved in binding other proteins, peptidomimetic compounds can be generated which mimic those residues which facilitate the interaction. Such mimetics may then be used to interfere with the normal function of a Protease M protein. For instance, non-hydrolyzable peptide analogs of such residues can be generated using benzodiazepine (e.g., see Freidinger et al. in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), azepine (e.g., see Huffman et al. in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), substituted gamma lactam rings (Garvey et al. in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands

1988), keto-methylene pseudopeptides (Ewenson et al. (1986) *J Med Chem* 29:295; and Ewenson et al. in Peptides: Structure and Function (Proceedings of the 9th American Peptide Symposium) Pierce Chemical Co. Rockland, IL, 1985), b-turn dipeptide cores (Nagai et al. (1985) *Tetrahedron Lett* 26:647; and Sato et al. (1986) *J Chem Soc* Perkin Trans 1:1231), and b-aminoalcohols (Gordon et al. (1985) *Biochem Biophys Res Commun* 126:419; and Dann et al. (1986) *Biochem Biophys Res Commun* 134:71).

V. Antibodies

Another aspect of the invention pertains to anti-Protease M antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as Protease M. The invention provides polyclonal and monoclonal antibodies that bind Protease M. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of Protease M. A monoclonal antibody composition thus typically displays a single binding affinity for a particular Protease M protein with which it immunoreacts.

Polyclonal Protease M antibodies can be prepared as described above by immunizing a suitable subject with a Protease M immunogen, as described in more detail in the appended Examples. The anti-Protease M antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized Protease M. If desired, the antibody molecules directed against Protease M can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, e.g., when the anti-Protease M antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975, *Nature* 256:495-497) (see also, Brown et al. (1981) *J. Immunol* 127:539-46; Brown et al. (1980) *J Biol Chem* 255:4980-83; Yeh et al. (1976) *PNAS* 76:2927-31; and Yeh et al. (1982) *Int. J. Cancer* 29:269-75), the more recent human B cell hybridoma technique (Kozbor et al. (1983) *Immunol Today* 4:72), the EBV-hybridoma technique (Cole et al. (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing monoclonal antibody hybridomas is well known (see generally R. H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New

York, New York (1980); E. A. Lerner (1981) *Yale J. Biol. Med.*, 54:387-402; M. L. Gefter et al. (1977) *Somatic Cell Genet.*, 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a Protease M immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds Protease M.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-Protease M monoclonal antibody (see, e.g., G. Galfre et al. (1977) *Nature* 266:55052; Gefter et al. *Somatic Cell Genet.*, cited *supra*; Lerner, *Yale J. Biol. Med.*, cited *supra*; Kenneth, *Monoclonal Antibodies*, cited *supra*). Moreover, the person of ordinary skill in the art will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (e.g., a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line. Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines may be used as a fusion partner according to standard techniques, e.g., the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from the American Type Culture Collection (ATCC), Rockville, Md. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days because they are not transformed).

Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind Protease M, e.g., using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-Protease M antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with Protease M to thereby isolate immunoglobulin library members that bind Protease M. Kits for generating and screening phage display libraries are commercially available (e.g., the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP*™ *Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for

generating and screening antibody display library can be found in, for example, Ladner et al. U.S. Patent No. 5,223,409; Kang et al. International Publication No. WO 92/18619; Dower et al. International Publication No. WO 91/17271; Winter et al. International Publication WO 92/20791; Markland et al. International Publication No. WO 92/15679; Breitling et al. International Publication WO 93/01288; McCafferty et al. International Publication No. WO 92/01047; Garrard et al. International Publication No. WO 92/09690; Ladner et al. International Publication No. WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-1372; Hay et al. (1992) *Hum Antibod Hybridomas* 3:81-85; Huse et al. (1989) *Science* 246:1275-1281; Griffiths et al. (1993) *EMBO J* 12:725-734; Hawkins et al. (1992) *J Mol Biol* 226:889-896; Clarkson et al. (1991) *Nature* 352:624-628; Gram et al. (1992) *PNAS* 89:3576-3580; Garrad et al. (1991) *Bio/Technology* 9:1373-1377; Hoogenboom et al. (1991) *Nuc Acid Res* 19:4133-4137; Barbas et al. (1991) *PNAS* 88:7978-7982; and McCafferty et al. *Nature* (1990) 348:552-554.

Additionally, recombinant anti-Protease M antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson et al. International Patent Publication PCT/US86/02269; Akira, et al. European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison et al. European Patent Application 173,494; Neuberger et al. PCT Application WO 86/01533; Cabilly et al. U.S. Patent No. 4,816,567; Cabilly et al. European Patent Application 125,023; Better et al. (1988) *Science* 240:1041-1043; Liu et al. (1987) *PNAS* 84:3439-3443; Liu et al. (1987) *J. Immunol.* 139:3521-3526; Sun et al. (1987) *PNAS* 84:214-218; Nishimura et al. (1987) *Canc. Res.* 47:999-1005; Wood et al. (1985) *Nature* 314:446-449; and Shaw et al. (1988) *J. Natl Cancer Inst.* 80:1553-1559); Morrison, S. L. (1985) *Science* 229:1202-1207; Oi et al. (1986) *BioTechniques* 4:214; Winter U.S. Patent 5,225,539; Jones et al. (1986) *Nature* 321:552-525; Verhoeyan et al. (1988) *Science* 239:1534; and Beidler et al. (1988) *J. Immunol.* 141:4053-4060.

An anti-Protease M antibody (e.g., monoclonal antibody) can be used to isolate Protease M by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-Protease M antibody can facilitate the purification of natural Protease M from cells and of recombinantly produced Protease M expressed in host cells. Moreover, an anti-Protease M antibody can be used to detect Protease M protein (e.g., in a cellular lysate or cell supernatant). Detection may be facilitated by

coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

VI. Transgenic animals

Another aspect of the invention features transgenic non-human animals which express a heterologous Protease M gene of the present invention, or which have had one or more genomic Protease M genes disrupted in at least one of the tissue or cell-types of the animal. Accordingly, the invention features an animal model for proliferative disorders, which animal has one or more Protease M allele which is mis-expressed. For example, a mouse can be bred which has one or more Protease M alleles deleted or otherwise rendered inactive. Such a mouse model can then be used to study disorders arising from mis-expressed Protease M genes, as well as for evaluating potential therapies for similar disorders.

Another aspect of the present invention concerns transgenic animals which are comprised of cells (of that animal) which contain a transgene of the present invention and which preferably (though optionally) express an exogenous Protease M protein in one or more cells in the animal. A Protease M transgene can encode the wild-type form of the protein, or can encode homologs thereof, including both agonists and antagonists, as well as antisense constructs. In preferred embodiments, the expression of the transgene is restricted to specific subsets of cells, tissues or developmental stages utilizing, for example, cis-acting sequences that control expression in the desired pattern. In the present invention, such mosaic expression of a Protease M protein can be essential for many forms of lineage analysis and can additionally provide a means to assess the effects of, for example, lack of Protease M expression which might grossly alter development in small patches of tissue within an otherwise normal embryo. Toward this end, tissue-specific regulatory sequences and conditional regulatory sequences can be used to control expression of the transgene in certain spatial patterns. Moreover, temporal patterns of expression can be provided by, for example, conditional recombination systems or prokaryotic transcriptional regulatory sequences.

Genetic techniques which allow for the expression of transgenes can be regulated via site-specific genetic manipulation *in vivo* are known to those skilled in the art. For instance, genetic systems are available which allow for the regulated expression of a recombinase that catalyzes the genetic recombination a target sequence. As used herein, the phrase "target sequence" refers to a nucleotide sequence that is genetically recombined by a recombinase. The target sequence is flanked by recombinase recognition sequences and is generally either excised or inverted in cells expressing recombinase activity. Recombinase catalyzed recombination events can be designed such that recombination of the target sequence results in either the activation or repression of expression of one of the subject Protease M proteins. For example, excision of a target sequence which interferes with the expression of a recombinant Protease M gene, such as one which encodes an antagonistic homolog or an antisense transcript, can be designed to activate expression of that gene. This interference with expression of the protein can result from a variety of mechanisms, such as spatial separation of the Protease M gene from the promoter element or an internal stop codon. Moreover, the transgene can be made wherein the coding sequence of the gene is flanked by recombinase recognition sequences and is initially transfected into cells in a 3' to 5' orientation with respect to the promoter element. In such an instance, inversion of the target sequence will reorient the subject gene by placing the 5' end of the coding sequence in an orientation with respect to the promoter element which allow for promoter driven transcriptional activation.

In an illustrative embodiment, either the cre/loxP recombinase system of bacteriophage P1 (Lakso et al. (1992) PNAS 89:6232-6236; Orban et al. (1992) PNAS 89:6861-6865) or the FLP recombinase system of Saccharomyces cerevisiae (O'Gorman et al. (1991) Science 251:1351-1355; PCT publication WO 92/15694) can be used to generate *in vivo* site-specific genetic recombination systems. Cre recombinase catalyzes the site-specific recombination of an intervening target sequence located between loxP sequences. loxP sequences are 34 base pair nucleotide repeat sequences to which the Cre recombinase binds and are required for Cre recombinase mediated genetic recombination. The orientation of loxP sequences determines whether the intervening target sequence is excised or inverted when Cre recombinase is present (Abremski et al. (1984) J. Biol. Chem. 259:1509-1514); catalyzing the excision of the target sequence when the loxP sequences are oriented as direct repeats and catalyzes inversion of the target sequence when loxP sequences are oriented as inverted repeats.

Accordingly, genetic recombination of the target sequence is dependent on expression of the Cre recombinase. Expression of the recombinase can be regulated by promoter elements which are subject to regulatory control, e.g., tissue-specific,

developmental stage-specific, inducible or repressible by externally added agents. This regulated control will result in genetic recombination of the target sequence only in cells where recombinase expression is mediated by the promoter element. Thus, the activation expression of a recombinant Protease M protein can be regulated via control of recombinase expression.

Use of the cre/loxP recombinase system to regulate expression of a recombinant Protease M protein requires the construction of a transgenic animal containing transgenes encoding both the Cre recombinase and the subject protein. Animals containing both the Cre recombinase and a recombinant Protease M gene can be provided through the construction of "double" transgenic animals. A convenient method for providing such animals is to mate two transgenic animals each containing a transgene, e.g., a Protease M gene and recombinase gene.

One advantage derived from initially constructing transgenic animals containing a Protease M transgene in a recombinase-mediated expressible format derives from the likelihood that the subject protein, whether agonistic or antagonistic, can be deleterious upon expression in the transgenic animal. In such an instance, a founder population, in which the subject transgene is silent in all tissues, can be propagated and maintained. Individuals of this founder population can be crossed with animals expressing the recombinase in, for example, one or more tissues and/or a desired temporal pattern. Thus, the creation of a founder population in which, for example, an antagonistic Protease M transgene is silent will allow the study of progeny from that founder in which disruption of Protease M mediated induction in a particular tissue or at certain developmental stages would result in, for example, a lethal phenotype.

Similar conditional transgenes can be provided using prokaryotic promoter sequences which require prokaryotic proteins to be simultaneous expressed in order to facilitate expression of the Protease M transgene. Exemplary promoters and the corresponding trans-activating prokaryotic proteins are given in U.S. Patent No. 4,833,080.

Moreover, expression of the conditional transgenes can be induced by gene therapy-like methods wherein a gene encoding the trans-activating protein, e.g. a recombinase or a prokaryotic protein, is delivered to the tissue and caused to be expressed, such as in a cell-type specific manner. By this method, a Protease M transgene could remain silent into adulthood until "turned on" by the introduction of the trans-activator.

In an exemplary embodiment, the "transgenic non-human animals" of the invention are produced by introducing transgenes into the germline of the non-human animal. Embryonic target cells at various developmental stages can be used to intro-

transgenes. Different methods are used depending on the stage of development of the embryonic target cell. The zygote is the best target for micro-injection. In the mouse, the male pronucleus reaches the size of approximately 20 micrometers in diameter which allows reproducible injection of 1-2pl of DNA solution. The use of zygotes as a target for gene transfer has a major advantage in that in most cases the injected DNA will be incorporated into the host gene before the first cleavage (Brinster et al. (1985) PNAS 82:4438-4442). As a consequence, all cells of the transgenic non-human animal will carry the incorporated transgene. This will in general also be reflected in the efficient transmission of the transgene to offspring of the founder since 50% of the germ cells will harbor the transgene. Microinjection of zygotes is the preferred method for incorporating transgenes in practicing the invention.

Retroviral infection can also be used to introduce Protease M transgenes into a non-human animal. The developing non-human embryo can be cultured *in vitro* to the blastocyst stage. During this time, the blastomeres can be targets for retroviral infection (Jaenich, R. (1976) PNAS 73:1260-1264). Efficient infection of the blastomeres is obtained by enzymatic treatment to remove the zona pellucida (Manipulating the Mouse Embryo, Hogan eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1986). The viral vector system used to introduce the transgene is typically a replication-defective retrovirus carrying the transgene (Jahner et al. (1985) PNAS 82:6927-6931; Van der Putten et al. (1985) PNAS 82:6148-6152). Transfection is easily and efficiently obtained by culturing the blastomeres on a monolayer of virus-producing cells (Van der Putten, supra; Stewart et al. (1987) EMBO J. 6:383-388). Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can be injected into the blastocoele (Jahner et al. (1982) Nature 298:623-628). Most of the founders will be mosaic for the transgene since incorporation occurs only in a subset of the cells which formed the transgenic non-human animal. Further, the founder may contain various retroviral insertions of the transgene at different positions in the genome which generally will segregate in the offspring. In addition, it is also possible to introduce transgenes into the germ line by intrauterine retroviral infection of the midgestation embryo (Jahner et al. (1982) supra).

A third type of target cell for transgene introduction is the embryonic stem cell (ES). ES cells are obtained from pre-implantation embryos cultured *in vitro* and fused with embryos (Evans et al. (1981) Nature 292:154-156; Bradley et al. (1984) Nature 309:255-258; Gossler et al. (1986) PNAS 83: 9065-9069; and Robertson et al. (1986) Nature 322:445-448). Transgenes can be efficiently introduced into the ES cells by DNA transfection or by retrovirus-mediated transduction. Such transformed ES cells can thereafter be combined with blastocysts from a non-human animal. The ES cells

thereafter colonize the embryo and contribute to the germ line of the resulting chimeric animal. For review see Jaenisch, R. (1988) Science 240:1468-1474.

Methods of making Protease M knock-out or disruption transgenic animals are also generally known. See, for example, Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Recombinase dependent knockouts can also be generated, e.g. by homologous recombination to insert recombinase target sequences flanking portions of an endogenous Protease M gene, such that tissue specific and/or temporal control of inactivation of a Protease M allele can be controlled as above.

VII. Pharmaceutical Compositions

The Protease M proteins, Protease M nucleic acids, and anti-Protease M antibodies of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the protein or antibody and a pharmaceutically acceptable carrier. As used herein the term "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. For example, solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For

intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a Protease M protein or anti-Protease M antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or

lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These may be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on (a) the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and (b) the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

VIII. Uses and Methods of the Invention

As described in more detail in the appended Examples, the Protease M protein of the invention exhibits serine protease activity. Accordingly, Protease M is useful as a serine protease, either *in vitro* or *in vivo*. The isolated nucleic acid molecules of the invention can be used to express Protease M protein (e.g., via a recombinant expression vector in a host cell), to detect Protease M mRNA (e.g., in a biological sample) and to modulate Protease M activity, as discussed and further below. Moreover, the anti-Protease M antibodies of the invention can be used to detect and isolate Protease M protein and modulate Protease M activity, also discussed further below.

A. Diagnostic and Prognostic Assays

The present method provides a method for determining if a subject is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In preferred embodiments, the methods can be characterized as comprising detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of (i) an alteration affecting the integrity of a gene encoding a *Protease M*-protein, or (ii) the mis-expression of the *Protease M* gene. To illustrate, such genetic lesions can be detected by ascertaining the existence of at least one of (i) a deletion of one or more nucleotides from a *Protease M* gene, (ii) an addition of one or more nucleotides to a *Protease M* gene, (iii) a substitution of one or more nucleotides of a *Protease M* gene, (iv) a gross chromosomal rearrangement of a *Protease M* gene, (v) a gross alteration in the level of a messenger RNA transcript of a *Protease M* gene, (vii) aberrant modification of a *Protease M* gene, such as of the methylation pattern of the genomic DNA, (vii) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a *Protease M* gene, (viii) a non-wild type level of a *Protease M*-protein, (ix) allelic loss of a *Protease M* gene, and (x) inappropriate post-translational modification of a *Protease M*-protein. As set out below, the present invention provides a large number of assay techniques for detecting lesions in a *Protease M* gene, and importantly, provides the ability to discern between different molecular causes underlying *Protease M*-dependent aberrant cell growth, proliferation and/or differentiation.

In an exemplary embodiment, there is provided a nucleic acid composition comprising a (purified) oligonucleotide probe including a region of nucleotide sequence which is capable of hybridizing to a sense or antisense sequence of a *Protease M* gene, such as represented by any of SEQ ID Nos: 1 and 3, or naturally occurring mutants thereof, or 5' or 3' flanking sequences or intronic sequences naturally associated with the subject *Protease M* genes or naturally occurring mutants thereof. The nucleic acid of a cell is rendered accessible for hybridization, the probe is contacted with nucleic acid of the sample, and the hybridization of the probe to the sample nucleic acid is detected. Such techniques can be used to detect lesions at either the genomic or mRNA level, including deletions, substitutions, etc., as well as to determine mRNA transcript levels.

As set out above, one aspect of the present invention relates to diagnostic assays for determining, in the context of cells isolated from a patient, if mutations have arisen in one or more *Protease M* of the sample cells. The present method provides a method for determining if a subject is at risk for a disorder characterized by aberrant cell proliferation and/or metastasis. In preferred embodiments, the method can be generally

characterized as comprising detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by an alteration affecting the integrity of a gene encoding a *Protease M*. To illustrate, such genetic lesions can be detected by ascertaining the existence of at least one of (i) a deletion of one or more nucleotides from a *Protease M*-gene, (ii) an addition of one or more nucleotides to a *Protease M*-gene, (iii) a substitution of one or more nucleotides of a *Protease M*-gene, and (iv) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a *Protease M*-gene. As set out below, the present invention provides a large number of assay techniques for detecting lesions in *Protease M* genes, and importantly, provides the ability to discern between different molecular causes underlying *Protease M*-dependent aberrant cell growth and/or metastasis.

In certain embodiments, detection of the lesion comprises utilizing the probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran et al. (1988) *Science* 241:1077-1080; and Nakazawa et al. (1994) *PNAS* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the *Protease M*-gene (see Abravaya et al. (1995) *Nuc Acid Res* 23:675-682). In a merely illustrative embodiment, the method includes the steps of (i) collecting a sample of cells from a patient, (ii) isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, (iii) contacting the nucleic acid sample with one or more primers which specifically hybridize to a *Protease M* gene under conditions such that hybridization and amplification of the *Protease M*-gene (if present) occurs, and (iv) detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein. Alternative amplification methods include: self sustained sequence replication (Guatelli, J.C. et al., 1990, Proc. Natl. Acad. Sci. USA 87:1874-1878), transcriptional amplification system (Kwoh, D.Y. et al., 1989, Proc. Natl. Acad. Sci. USA 86:1173-1177), Q-Beta Replicase (Lizardi, P.M. et al., 1988, Bio/Technology 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In another embodiment of the subject assay, mutations in a *Protease M* gene from a sample cell are identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally),

digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the *Protease M* gene and detect mutations by comparing the sequence of the sample *Protease M* with the corresponding wild-type (control) sequence. Exemplary sequencing reactions include those based on techniques developed by Maxim and Gilbert (*Proc. Natl Acad Sci USA* (1977) 74:560) or Sanger (Sanger et al (1977) *Proc. Nat. Acad. Sci* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures may be utilized when performing the subject assays (*Biotechniques* (1995) 19:448), including by sequencing by mass spectrometry (see, for example PCT publication WO 94/16101; Cohen et al. (1996) *Adv Chromatogr* 36:127-162; and Griffin et al. (1993) *Appl Biochem Biotechnol* 38:147-159). It will be evident to one skilled in the art that, for certain embodiments, the occurrence of only one, two or three of the nucleic acid bases need be determined in the sequencing reaction. For instance, A-tract or the like, e.g., where only one nucleic acid is detected, can be carried out.

In a further embodiment, protection from cleavage agents (such as a nuclease, hydroxylamine or osmium tetroxide and with piperidine) can be used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers, et al. (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type *Protease M* sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton et al (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba et al (1992) Methods Enzymod. 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in *Protease M* cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu et al. (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on a *Protease M* sequence, e.g., a wild-type *Protease M* sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in *Protease M* genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita et al. (1989) *Proc Natl. Acad. Sci USA* 86:2766, see also Cotton (1993) *Mutat Res* 285:125-144; and Hayashi (1992) *Genet Anal Tech Appl* 9:73-79). Single-stranded DNA fragments of sample and control *Protease M* nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen et al. (1991) Trends Genet 7:5).

In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers et al (1985) Nature 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing agent gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) Biophys Chem 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or

selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki et al. (1986) *Nature* 324:163); Saiki et al (1989) *Proc. Natl Acad. Sci USA* 86:6230). Such allele speicific oligonucleotide hybridization techniques may be used to test one mutation per reaction when oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs et al (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238. In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini et al (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

Another embodiment of the invention provides for a nucleic acid composition comprising a (purified) oligonucleotide probe including a region of nucleotide sequence which is capable of hybridizing to a sense or antisense sequence of a *Protease M*-gene, or naturally occurring mutants thereof, or 5' or 3' flanking sequences or intronic sequences naturally associated with the subject *Protease M*-genes or naturally occurring mutants thereof. The nucleic acid of a cell is rendered accessible for hybridization, the probe is exposed to nucleic acid of the sample, and the hybridization of the probe to the sample nucleic acid is detected. Such techniques can be used to detect lesions at either the genomic or mRNA level, including deletions, substitutions, etc., as well as to determine mRNA transcript levels. Such oligonucleotide probes can be used for both predictive and therapeutic evaluation of allelic mutations which might be manifest in, for example, neoplastic or hyperplastic disorders (e.g. aberrant cell growth).

To illustrate, nucleotide probes can be generated from the subject Protease M gene which facilitate histological screening of intact tissue and tissue

samples for the presence (or absence) of Protease M-encoding transcripts. Similar to the diagnostic uses of anti-Protease M antibodies, the use of probes directed to Protease M messages, or to genomic Protease M sequences, can be used for both predictive and therapeutic evaluation of allelic mutations which might be manifest in, for example, neoplastic or hyperplastic disorders (e.g. unwanted cell growth) or abnormal differentiation of tissue. Used in conjunction with immunoassays as described above, the oligonucleotide probes can help facilitate the determination of the molecular basis for a developmental disorder which may involve some abnormality associated with expression (or lack thereof) of a Protease M protein. For instance, variation in polypeptide synthesis can be differentiated from a mutation in a coding sequence.

Diagnostic procedures may be performed on any "biological sample" including, for example, cells, body fluids, or *in situ* directly upon tissue sections (fixed and/or frozen) of patient tissue obtained from biopsies or resections, such that no nucleic acid purification is necessary.

Antibodies directed against wild type or mutant *Protease M* proteins, which are discussed, above, may also be used in disease diagnostics and prognostics. Such diagnostic methods, may be used to detect abnormalities in the level of *Protease M* protein expression, or abnormalities in the structure and/or tissue, cellular, or subcellular location of *Protease M* protein. Structural differences may include, for example, differences in the size, electronegativity, or antigenicity of the mutant *Protease M* protein relative to the normal *Protease M* protein. Protein from the tissue or cell type to be analyzed may easily be detected or isolated using techniques which are well known to one of skill in the art, including but not limited to western blot analysis. For a detailed explanation of methods for carrying out western blot analysis, see Sambrook et al, 1989, supra, at Chapter 18. The protein detection and isolation methods employed herein may also be such as those described in Harlow and Lane, for example, (Harlow, E. and Lane, D., 1988, "Antibodies: A Laboratory Manual", Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York), which is incorporated herein by reference in its entirety.

This can be accomplished, for example, by any of a number of techniques known in the art, such as, for example, immunofluorescence techniques employing a fluorescently labeled antibody (see below) coupled with light microscopic, flow cytometric, or fluorimetric detection. The term "labeled or labelable", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that

may, additionally, be employed histologically, as in immunofluorescence or immunoelectron microscopy, for *in situ* detection of *Protease M* proteins. *In situ* detection may be accomplished by removing a histological specimen from a patient, and applying thereto a labeled antibody of the present invention. The antibody (or fragment) is preferably applied by overlaying the labeled antibody (or fragment) onto a biological sample. Through the use of such a procedure, it is possible to determine not only the presence of the *Protease M* protein, but also its distribution in the examined tissue. Using the present invention, one of ordinary skill will readily perceive that any of a wide variety of histological methods (such as staining procedures) can be modified in order to achieve such *in situ* detection.

One means for labeling an anti- *Protease M* protein specific antibody is via linkage to an enzyme and use in an enzyme immunoassay (EIA) (Voller, "The Enzyme Linked Immunosorbent Assay (ELISA)", *Diagnostic Horizons* 2:1-7, 1978, Microbiological Associates Quarterly Publication, Walkersville, MD; Voller, et al., J. Clin. Pathol. 31:507-520 (1978); Butler, Meth. Enzymol. 73:482-523 (1981); Maggio, (ed.) *Enzyme Immunoassay*, CRC Press, Boca Raton, FL, 1980; Ishikawa, et al., (eds.) *Enzyme Immunoassay*, Kgaku Shoin, Tokyo, 1981). The enzyme which is bound to the antibody will react with an appropriate substrate, preferably a chromogenic substrate, in such a manner as to produce a chemical moiety which can be detected, for example, by spectrophotometric, fluorimetric or by visual means. Enzymes which can be used to detectably label the antibody include, but are not limited to, malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast alcohol dehydrogenase, alpha-glycerophosphate, dehydrogenase, triose phosphate isomerase, horseradish peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase and acetylcholinesterase. The detection can be accomplished by colorimetric methods which employ a chromogenic substrate for the enzyme. Detection may also be accomplished by visual comparison of the extent of enzymatic reaction of a substrate in comparison with similarly prepared standards.

Detection may also be accomplished using any of a variety of other methods. Antibodies may be labeled with radioactivity, fluorescent compounds (e.g., fluorescein isothiocyanate, rhodamine, phycoerythrin, phycocyanin, allophycocyanin, o-phthaldehyde and fluorescamine), chemiluminescent compounds (e.g., luminol, isoluminol, theromatic acridinium ester, imidazole, acridinium salt and oxalate ester), bioluminescent compounds (e.g., luciferin, luciferase and aequorin).

Moreover, it will be understood that any of the above methods for detecting alterations in a *Protease M* gene or gene product can be used to monitor the course of treatment or therapy.

In another embodiment detection of Protease M is based on the detection a Protease M bioactivity, such as enzymatic activity. For example, serine protease substrate cleavage may be measured in a sample. Exemplary substrates include, gelatin, casein, or n-a-benzoyl-L-arginine ethyl ester (BAEE).

In an exemplary embodiment, the invention provides a diagnostic method comprising: (i) contacting a tumor sample from a subject with an agent capable of detecting Protease M protein or mRNA; (2) determining the amount of Protease M protein or mRNA expressed in the tumor sample; (3) comparing the amount of Protease M protein or mRNA expressed in the tumor sample to a control sample; and (4)forming a diagnosis based on the amount of Protease M protein or mRNA expressed in the tumor sample as compared to the control sample.

In a preferred embodiment of the detection method, the biological sample is a tumor sample. The tumor sample may comprise tumor tissue or a suspension of tumor cells. A tissue section, for example, a freeze-dried or fresh frozen section of tumor tissue removed from a patient, can be used as the tumor sample. Moreover, the tumor sample may comprise a biological fluid obtained from a tumor-bearing subject. Protease M contains a signal sequence and thus is likely to be detectable in biological fluids. Following collection, tumor samples can be stored at temperatures below -20^{o}C to prevent degradation until the detection method is to be performed. Preferred tumor samples in which Protease M mRNA or protein is to be detected are mammary tumor samples and/or ovarian tumor samples. Primary malignancy of the tumor cell sample can be diagnosed based on an increase in the level of expression of Protease M mRNA or protein in the tumor sample as compared to the control. In another embodiment, the control is from normal cells or a primary tumor and the tumor sample is a suspected metastatic tumor sample. Acquisition of the metastatic phenotype by the suspected metastatic tumor sample can be diagnosed based on a decrease in the level of, or absence of, Protease M mRNA or protein in the tumor sample compared to the control. The detection method of the invention can be used to detect Protease M mRNA or protein in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of Protease M mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of Protease M protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. Alternatively, Protease M protein can be detected *in vivo* in a subject by introducing into the subject a labeled anti-Protease M antibody. For example

the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a *Protease M* gene. For example, the kit can comprise a labeled or labelable agent capable of detecting Protease M protein or mRNA in a biological sample; means for determining the amount of Protease M in the sample; and means for comparing the amount of Protease M in the sample with a standard. The agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect Protease M mRNA or protein.

B. Therapeutic Uses

Another aspect of the invention pertains to methods of modulating Protease M bioactivity associated with a cell, e.g., for therapeutic purposes. Protease M activity "associated with a cell" is intended to include Protease M activity within the cell, secreted by the cell and in the extracellular milieu surrounding the cell. The modulatory method of the invention involves contacting the cell with an agent that modulates Protease M activity associated with the cell. In one embodiment, the agent stimulates Protease M serine protease activity. Examples of such stimulatory agents include active Protease M protein agonists and a nucleic acid molecule encoding Protease M that has been introduced into the cell. In another embodiment, the agent inhibits the Protease M activity . Examples of such inhibitory agents include antisense Protease M nucleic acid molecules, anti-Protease M antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject).

Stimulation of Protease M bioactivity is desirable in situations in which Protease M is abnormally downregulated and/or in which increased Protease M activity is likely to have a beneficial effect. One example of such a situation is in tumor cells, and in particular metastatic tumor cells. As demonstrated in the appended Examples, acquisition of a metastatic phenotype by tumor cells is associated with downregulation of Protease M expression. Thus, increasing the expression and/or activity of Protease M in or around the tumor cells is expected to inhibit the development or progression of the metastatic phenotype. Accordingly, in a specific embodiment, the invention provides a method for inhibiting development or progression of a tumorogeinc phenotype in a cell comprising contacting the tumor cell with an agent which elevates the amount of

Protease M associated with the cell. The agent that elevates Protease M in or around the tumor cell can be Protease M protein itself. For example, since Protease M is likely to be a secreted protein, it is likely that it exerts tumor suppressive effects extracellularly. Thus, Protease M, preferably in a pharmaceutically acceptable carrier, can be administered to a tumor-bearing subject by an appropriate route to inhibit the development or progression of a proliferative disorder. Suitable routes of administration include intravenous, intramuscular or subcutaneous injection, injection directly into the tumor site or implantation of a device containing a slow-release formulation. The Protease M preparation can also be incorporated into liposomes or other carrier vehicles to facilitate delivery to the tumor site. A non-limiting dosage range is 0.001 to 100 mg/kg/day, with the most beneficial range to be determined by routine pharmacological methods.

Alternative to administration of Protease M protein or agonist itself, the development of or progression of cancer in a cell may be slowed by modifying them to express Protease M by introducing into the cells a nucleic acid encoding Protease M (e.g., via a recombinant expression vector). Expression vectors suitable for gene therapy, including retroviral and adenoviral vectors carrying appropriate regulatory elements, can be used to deliver the Protease M-encoding nucleic acid to the tumor cells.

The ability of Protease M protein or DNA to inhibit tumor progression and/or metastasis can be evaluated using *in vivo* and *in vitro* assays known in the art. For example, a suitable *in vivo* assay utilizes the mammary epithelial tumor cell line MDA-MB-435, which forms tumors at the site of orthotopic injection and metastasizes in nude mice (describe further in Price et al. (1990) *Cancer Res.* 50:717). MDA-MB-435 cells, which do not express detectable Protease M mRNA, can be transfected with a Protease M expression vector and stable transfectants can be selected. These transfectants can then be injected into nude mice. At 10-weeks post-inoculation, the mice are sacrificed and their tumors are excised and weighed to determine the effect of Protease M expression on tumor progression and metastasis. A suitable in vitro assay is tumor cell invasion through reconstituted basement membrane matrix (e.g., Matrigel) as described in Hendrix et al. (1987) Cancer Letters 38:137. The invasive ability of Protease M-transfected MDA-MB-435 cells can be compared to untransfected MDA-MB-435 cells to determine the effect of Protease M expression on tumor invasiveness.

In contrast to the foregoing situations in which stimulation of Protease M activity is desirable, there are other situations in which it may be desirable to decrease Protease M activity using an inhibitory method of the invention. For example, as demonstrated herein, Protease M mRNA expression is markedly upregulated in certain

primary tumor cells. Thus, inhibiting the expression or activity of Protease M in cells may be useful for inhibiting or reducing carcinogenesis.

C. Drug Screening Assays

Furthermore, by making available purified and recombinant *Protease M* polypeptides, the present invention facilitates the development of assays which can be used to screen for drugs, including *Protease M* homologs, which are either agonists or antagonists of the normal cellular function of the subject *Protease M* polypeptides, or of their role in the pathogenesis of cellular differentiation and/or proliferation and disorders related thereto. In one embodiment, the assay evaluates the ability of a compound to modulate binding between a *Protease M* polypeptide and a molecule, be it protein or DNA, that interacts either upstream or downstream of the *Protease M* polypeptide in the TGFb signaling pathway. A variety of assay formats will suffice and, in light of the present inventions, will be comprehended by a skilled artisan.

1. Cell-free assays

In many drug screening programs which test libraries of compounds and natural extracts, high throughput assays are desirable in order to maximize the number of compounds surveyed in a given period of time. Assays which are performed in cell-free systems, such as may be derived with purified or semi-purified proteins, are often preferred as "primary" screens in that they can be generated to permit rapid development and relatively easy detection of an alteration in a molecular target which is mediated by a test compound. Moreover, the effects of cellular toxicity and/or bioavailability of the test compound can be generally ignored in the *in vitro* system, the assay instead being focused primarily on the effect of the drug on the molecular target as may be manifest in an alteration of binding affinity with upstream or downstream elements.

In an exemplary screening assay of the present invention, the compound of interest is contacted with Protease M and a molecule which interacts with Protease M (including both activators and repressors of its activity), such as a substrate. Detection and quantification of complexes of *Protease M* with it's binding protein provide a means for determining a compound's efficacy at inhibiting (or potentiating) complex formation between *Protease M* and the *Protease M*-binding elements. The efficacy of the compound can be assessed by generating dose response curves from data obtained using various concentrations of the test compound. Moreover, a control assay can also be performed to provide a baseline for comparison. In the control assay, isolated and purified *Protease M* polypeptide is added to a composition containing the *Protease M*-

binding element, and the formation of a complex is quantitated in the absence of the test compound.

Complex formation between the *Protease M* polypeptide and a *Protease M* binding element may be detected by a variety of techniques. Modulation of the formation of complexes can be quantitated using, for example, detectably labeled proteins such as radiolabeled, fluorescently labeled, or enzymatically labeled *Protease M* polypeptides, by immunoassay, or by chromatographic detection.

For example, modulators of Protease M activity may be identified in a method wherein Protease M, a substrate for the serine protease, and a test substance are incubated under conditions suitable for the serine protease to cleave the substrate. Cleavage of the substrate is then measured and the amount of cleavage of the substrate in the presence of the test substance is compared to the amount of cleavage of the substrate in the absence of the test substance. The test substance can then be identified as a modulator of Protease M activity based on this comparison. For example, when the amount of cleavage of the substrate in the presence of the test substance is less than the amount of cleavage of the substrate in the absence of the test substance, the test substance can thereby be identified as a stimulator of the Protease M activity. Alternatively, when the amount of cleavage of the substrate in the presence of the test substance is greater than the amount of cleavage of the substrate in the absence of the test substance, the test substance can thereby be identified as an inhibitor of the Protease M activity.

Typically, it will be desirable to immobilize either *Protease M* or its binding protein to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of *Protease M* to a binding protein, in the presence and absence of a candidate agent, can be accomplished in any vessel suitable for containing the reactants. Examples include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase/*Protease M* (GST/*Protease M*) fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates, e.g. an ^{35}S-labeled, and the test compound, and the mixture incubated under conditions conducive to complex formation, e.g. at physiological conditions for salt and pH, though slightly more stringent conditions may be desired. Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly (e.g. beads placed in scintilant), or in the supernatant after the complexes are subsequently dissociated. Alternatively, the complexes can be

dissociated from the matrix, separated by SDS-PAGE, and the level of *Protease M*-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques such as described in the appended examples.

Other techniques for immobilizing proteins on matrices are also available for use in the subject assay. For instance, either *Protease M* or its cognate binding protein can be immobilized utilizing conjugation of biotin and streptavidin. For instance, biotinylated *Protease M* molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with *Protease M* but which do not interfere with binding of Protease M and a binding protein (BP) can be derivatized to the wells of the plate, and *Protease M* trapped in the wells by antibody conjugation. As above, preparations of a *Protease M*-binding protein and a test compound are incubated in the *Protease M*-presenting wells of the plate, and the amount of complex trapped in the well can be quantitated. Exemplary methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the *Protease M* binding element, or which are reactive with *Protease M* protein and compete with the binding element; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the binding element, either intrinsic or extrinsic activity. In the instance of the latter, the enzyme can be chemically conjugated or provided as a fusion protein with the *Protease M*-BP. To illustrate, the *Protease M*-BP can be chemically cross-linked or genetically fused with horseradish peroxidase, and the amount of polypeptide trapped in the complex can be assessed with a chromogenic substrate of the enzyme, e.g. 3,3'-diamino-benzadine terahydrochloride or 4-chloro-1-napthol. Likewise, a fusion protein comprising the polypeptide and glutathione-S-transferase can be provided, and complex formation quantitated by detecting the GST activity using 1-chloro-2,4-dinitrobenzene (Habig et al (1974) J Biol Chem 249:7130).

For processes which rely on immunodetection for quantitating one of the proteins trapped in the complex, antibodies against the protein, such as anti-*Protease M* antibodies, can be used. Alternatively, the protein to be detected in the complex can be "epitope tagged" in the form of a fusion protein which includes, in addition to the *Protease M* sequence, a second polypeptide for which antibodies are readily available (e.g. from commercial sources). For instance, the GST fusion proteins described above can also be used for quantification of binding using antibodies against the GST moiety. Other useful epitope tags include myc-epitopes (e.g., see Ellison et al. (1991) J Biol Chem 266:21150-21157) which includes a 10-residue sequence from c-myc, as well as

the pFLAG system (International Biotechnologies, Inc.) or the pEZZ-protein A system (Pharamacia, NJ).

2. *Cell based assays*

In addition to cell-free assays, such as described above, the readily available source of mammalian *Protease M* proteins provided by the present invention also facilitates the generation of cell-based assays for identifying small molecule agonists/antagonists and the like. For example, cells can be caused to overexpress a recombinant *Protease M* protein in the presence and absence of a test agent of interest, and the assay scored for modulation in *Protease M* bioactivity in the target cell mediated by the test agent. As with the cell-free assays, agents which produce a statistically significant change in *Protease M*-dependent responses (either inhibition or potentiation) can be identified. In an illustrative embodiment, the expression or activity of a *Protease M* is modulated in or cells and the effects of compounds of interest on the readout of interest (such as tumorigenesis or metastatic potential) are measured.

In another embodiment, modulators of Protease M expression are identified in a method wherein a cell is contacted with a test substance and the expression of Protease M mRNA or protein in the cell is determined. The level of expression of Protease M mRNA or protein in the presence of the test substance is compared to the level of expression of Protease M mRNA or protein in the absence of the test substance. The test substance can then be identified as a modulator of Protease M expression based on this comparison. For example, when expression of Protease M mRNA or protein is greater in the presence of the test substance than in its absence, the test substance is identified as a stimulator of Protease M mRNA or protein expression. Alternatively, when expression of Protease M mRNA or protein is less in the presence of the test substance than in its absence, the test substance is identified as an inhibitor of Protease M mRNA or protein expression. The level of Protease M mRNA or protein expression in the cells can be determined by methods described above for detecting Protease M mRNA or protein. Alternatively, the regulatory regions of a Protease M gene, e.g., the 5' flanking promoter and enhancer regions, may be operably linked to a detectable marker (such as luciferase) which encodes a gene product that can be readily detected.

Monitoring the influence of compounds on cells may be applied not only in basic drug screening, but also in clinical trials. In such clinical trials, the expression of a panel of genes may be used as a "read out" of a particular drug's therapeutic effect.

In yet another aspect of the invention, the subject *Protease M* polypeptides can be used to generate a "two hybrid" assay (see, for example, U.S. Patent

No. 5,283,317; Zervos et al. (1993) Cell 72:223-232; Madura et al. (1993) J Biol Chem 268:12046-12054; Bartel et al. (1993) Biotechniques 14:920-924; Iwabuchi et al. (1993) Oncogene 8:1693-1696; and Brent WO94/10300), for isolating coding sequences for other cellular proteins which bind to or interact with *Protease M* ("*Protease M*-binding proteins" or "*Protease M*-bp"), such as a substrate. Such protease M-binding proteins would likely also be involved in the development of carcinogenesis or metastases.

Briefly, the two hybrid assay relies on reconstituting *in vivo* a functional transcriptional activator protein from two separate fusion proteins. In particular, the method makes use of chimeric genes which express hybrid proteins. To illustrate, a first hybrid gene comprises the coding sequence for a DNA-binding domain of a transcriptional activator fused in frame to the coding sequence for a *Protease M* polypeptide. The second hybrid protein encodes a transcriptional activation domain fused in frame to a sample gene from a cDNA library. If the bait and sample hybrid proteins are able to interact, e.g., form a *Protease M*-dependent complex, they bring into close proximity the two domains of the transcriptional activator. This proximity is sufficient to cause transcription of a reporter gene which is operably linked to a transcriptional regulatory site responsive to the transcriptional activator, and expression of the reporter gene can be detected and used to score for the interaction of the *Protease M* and sample proteins.

The present invention is further illustrated by the following examples which should not be construed as limiting in any way. The contents of all cited references (including literature references, issued patents, published patent applications as cited throughout this application are hereby expressly incorporated by reference.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, *Molecular Cloning A Laboratory Manual*, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989); *DNA Cloning*, Volumes I and II (D. N. Glover ed., 1985); *Oligonucleotide Synthesis* (M. J. Gait ed., 1984); Mullis et al. U.S. Patent No: 4,683,195; *Nucleic Acid Hybridization* (B. D. Hames & S. J. Higgins eds. 1984); *Transcription And Translation* (B. D. Hames & S. J. Higgins eds. 1984); *Culture Of Animal Cells* (R. I. Freshney, Alan R. Liss, Inc., 1987); *Immobilized Cells And Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide To Molecular Cloning* (1984); the treatise, *Methods In Enzymology* (Academic Press, Inc., N.Y.); *Gene*

Transfer Vectors For Mammalian Cells (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); *Methods In Enzymology*, Vols. 154 and 155 (Wu et al. eds.), *Immunochemical Methods In Cell And Molecular Biology* (Mayer and Walker, eds., Academic Press, London, 1987); *Handbook Of Experimental Immunology*, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

Exemplification

Example 1.

Materials and Methods Used in the Example.

Mammary Cell Strains and Lines

Normal human mammary epithelial cell strains (70N and 76N) were derived from reduction mammoplasties in this lab as described (Band V, Sager R. (1989) *Proc. Natl. Acad. Sci. USA86:1249-1253.*). Primary (21PT, 21NT) and metastatic (21MT-1, 21MT-2) tumor lines were established in this lab from a single patient as described (Band V and Sager R. (1 989) *Proc. Natl. Acad. Sci. USA86:1249-1253;* Band V, et al. (1990) *Cancer Res.* 50:735 1 -7357). Human mammary epithelial tumor cell lines MCF-7, T47D, ZR75-1, BT549, MDA-MB-157, MDA-MB-231, MDA-MB-435, MDA-MB436, MDA-MB-361, and BT-474 were obtained from American Tissue Culture Collection (Rockville, MD). Cells were grown in DFCI-1 media (Schachter M. (1980) *Pharmacol. Rev.* 31: 1 -1 7) and harvested at approximately 70% confluence for RNA isolation and when near confluent for DNA isolation.

Prostate Cell Lines

Normal, immortalized prostate epithelial cell lines: CF3 (HPV immortalized), CF91 (SV40 immortalized), and MLC (SV40 immortalized) were used in experiments. The tumor cell lines DU145, LNCaP, and PC3 (American Tissue Culture Collection, Rockville, MD) were also used.

Ovarian Cell Cultures and Tissues

The primary human ovarian surface epithelial cell cultures (HOSE 10/11, 16, and 21) were established from the ovarian surface epithelium as described previously (Tsao SW, Mok SC, Fey E, et al. (1995) *Exp. Cell Res.* 218:499-507). Immortalized ovarian surface epithelial cells (HOSE6.3E6E7) was obtained by infecting the HOSE cells with a replication defective retrovirus construct LXSN16E6E7 as described (Tsao

SW, et al. (1995). *Exp. Cell Res.* 218:499-507). The eight ovarian carcinoma cell lines used for this comparative study include DOV13, OVCA420, OVCA429, OVCA432, and OVCA433, which were established in the laboratory of Gynecologic Oncology; CAOV3 and SKOV3 which were purchased from ATCC (Rockville, MD); and OVCA3 which was obtained from the National Cancer Institute (Frederick, MD).

Ovarian tumors were obtained from consenting patients at the Brigham and Women's hospital in Boston as described previously (Mok SC, et al. (1992) *Cancer Res.* 52:5119-5122). These include six borderline ovarian tumors (354A, 373A, 395A, 405A, 466A, and 469A); twenty stage III/IV high grade invasive ovarian adenocarcinomas from the primary ovarian site; two metastatic adenocarcinoma from colon primary tumors (327A, 339A) and three normal ovaries (366N, 379N, and 465N).

Differential Display of mRNA

Total cell RNAs (50mg) from 21PT and 21MT-1 were treated with DNaseI (Worthington DPRF) in the presence of Rnasin ribonuclease inhibitor (Promega) to remove residual DNA contamination as described elsewhere (Sager R, et al. (1993). *FASEB J.* 7: 964-970). Differential display of the mRNA was performed as described (Liang L, Pardee AB. (1992) *Science* 257:967-970; Liang L, Averboukh L, Pardee AB. (1993) *Nucleic Acids Res.* 21:32673275). Basically the RNAs were reverse transcribed using the 3'-anchored primer T12MG (where M is a mixture of A, G, or C). The resultant cDNAs were then PCR amplified in the presence of 35S-dATP using T12MG and the arbitrary primer OPA1 (CAGGCCCTTC) and run side by side on a 6% sequencing gel. Differentially displayed bands were recovered from the dried gel, reamplified by PCR, ^{32}P-labeled by the oligo method (Feinberg AP and Vogelstein B. (1983) *Anal. Biochem.* 132:6-13) and used as a probe on Northern strips prepared with 21PT and 21MT-1 total RNA to confirm the result obtained by differential display.

Cloning and Sequencing of Partial and Full-Length cDNAs and Analysis

The reamplified band from differential display was cloned into the TA cloning vector PCRII (Invitrogen) and sequenced on both strands using T7 and SP6 primers. cDNA libraries from 21PT and 76N cells constructed in Lambda Zap II (Stratagene, San Diego, CA) were screened using the cloned PCR product as a probe and several cDNA clones were isolated and sequenced on both strands. The longest cDNA clone (from the 76N library) was sequenced on both strands using an ABI automated sequencer Model 373A by the Dana Farber Molecular Biology Core Facility. Oligonucleotides used for sequencing were synthesized by the Dana Farber Molecular Biology Core facility or by Amitof, Inc. (Cambridge, MA). The predicted protein

coding region and non-translated regions were determined and formatted using the GCG Publish program. The predicted protein sequence was compared to protein databases using the Blast algorithm (Altschul SF, (1990) *J. Mol. Biol.* 215:403-410). Protein alignment with related proteins performed on GCG using the Pileup, Distances, and Prettyplot programs.

Northern and Southern Analysis

Total cell RNA was isolated by the guanidinium isothiocyanate/cesium chloride method and analyzed on Northern blots as previously described (Sambrook J, (1989) Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.). 36B4 (Masiakowski P, Breathnach R, Bloch J, et al. (1982) *Nucl. Acids Res.* 10:7895-7903), a ribosomal protein whose message is constant under a variety of conditions, was used to normalize the blots. Total cellular DNA was isolated and analyzed on Southern blots as described (Sambrook J, (1989) Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.). Densitometric analysis of autoradiographs was performed with an imaging densitomer (Biorad GS-700) using the Molecular Analyst software.

Production of polyclonal antibody and western blotting

The MAP peptide (multiple antigen peptide) (Tam JP. (1988) *Proc. Natl. Acad. Sci.* USA 85:5409-5413.) ^{73}GKNNLRQRESSQEQS^{87} (0.5mg) was emulsified with an equal volume of Freund's adjuvant and injected into 3 to 9 month old New Zealand white rabbits. Boosts were done 2 and 6 weeks later. The animals were bled and serum was collected and stored at -20^{o} C. Peptide and antibody production was done at Research Genetics, Huntsville, AL.

Whole cell lysates were prepared by sonicating 10^7 cells/ml for 20, 30 second pulses in a Sonicator Ultrasonic Processor in mammalian lysis buffer (4mM$NaHCO_3$, 100mM NaF, 20mMKH_2PO_4, 2mM Sodium orthovanadate, 5mM EDTA, 5mM disoflurophoshate, 2mM PMSF, 2 mg/ ml leupeptin, 2mg/ml aprotinin, pH 7.2). Lysates were clarified by spinning at 14,000 x g for 30 minutes in a microfuge (Eppendorf).

50 to 100 mg of cell lysate was denatured by heating in SDS-PAGE sample buffer (50mM tris-HCl, pH6.8, 0.1mMDTT, 2% SDS, 0.1% bromphenol blue, 10% glycerol) at 90^{o}C for 5 minutes and run on a 12% premade acrylamide/ SDS minigel (biorad), electroblotted onto a PDVF membrane (0.2m, Biorad) and reacted with immune serum (1:1000). Anti-rabbit IgG horseradish peroxidase linked whole antibody

(Amersham) (1:2000) was used as secondary antibody, and immunoreactive bands were detected with the ECL (enhanced chemiluminescence, Amersham).

Expression of GST Fusion Protein

The full length cDNA clone was PCR amplified using the sense 5' 26-mer oligonucleotide $^{5'}$GGAATTCCGTTGGTGCATGGCGGACC$^{3'}$ and the antisense 3' oligonucleotide $^{5'}$GTCGGAATTCAGGGTCACTTGGCCTG$^{3'}$ at 95^{0}C, 1 minute, 60^{0}C, 1 minute, 72^{0}C, 1 minute for 30 cycles to yield a 0.7 kb product which contained the open reading frame without the hydrophobic n-terminal amino acids. The resultant PCR product encoding for leu^{22} to lys^{244} was digested with EcoR1 and ligated to alkaline phosphatase treated EcoRI linearized pGEX-2t vector (Pharmacia) to produce plasmids encoding a GST-Protease M fusion protein. E. Coli strains XL-1 blue or DH5a transformed with this construct were grown and induced with 0.2mM IPTG at 37^0 C for one hour to produce GST fusion protein which was solubilized from bacteria and purified on glutathionine agarose beads by standard methods (Smith DB, Johnson KS. (1988)*Gene* 67:31-40).

Expression of Baculovirus Recombinant Protein

A full length cDNA clone was cut with EcoN1 and BstX1 to give a fragment which spanned nucleotides 233 to 1019. This fragment was incorporated into the baculovirus transfer vector pVL1392 (Pharmingen). Generation and amplification of recombinant baculovirus was as described (23,24). For production of Protease M Spodoptera Frugiperda (cell line sf9) was infected with amplified recombinant virus to obtain nearly 100% infection as gauged by enlarged cells. 96 hours post-infection, cells were harvested and lysed by sonication in mammalian lysis buffer followed by adjusting to 500mM NaCl and rocking for one hour at 4^{o}C. All subsequent purifications were done at 4^{o}C.

The lysate was adjusted to 125mM NaCl, loaded onto p-aminobenzamidine agarose (Sigma A7155), washed with loading buffer, and eluted with (25mMNaPO$_4$, 0.02% NaN3, 500mM NaCl, 10mM benzamidine), pH 6.0. The eluted fractions were loaded onto concanavalin A agarose (Sigma C8402) by rocking for 1 hour, washed with (25mMNaPO$_4$, 0.02% NaN$_3$, 500 mM NaCl), pH 6.0, and eluted in wash buffer containing 10% methyl-a-D-mannopyranoside (Sigma M6882).

Assays for Protease Activity

Gelatin and casein zymography was performed essentially as described (Heusen C and Dowdle EB. (1980) *Anal. Biochem.* 102:196-202; Wilson MJ, et al..

(1993) *Journal of Urology* 149:65 3 -65 8.). Samples were run on 10% polyacryamide/0.1% SDS gels containing 1% gelatin or casein, soaked in 2.5% triton at room temperature for 1 hour, and in 0.1M glycine, pH 8.3 at 37°C. 5 to 16 hours. After staining in amido black areas of proteolysis appear as clear areas against the blue-black background. Trypsin (Sigma T8642) was used as a positive control.

Protease activity was also determined by monitoring the cleavage of n-a-benzoyl-L-arginine ethyl ester (BAEE) (Sigma B-4500. Reactions were set up in (25mMNaPO$_4$, 1mM EDTA, and 1mM BAEE), pH 7.5. Samples were added and the change in absorbance at 260nm was monitored on the Beckman DU-6 spectrophotometer in the time-drive mode. Trypsin was used as a positive control.

Expression Vector Construct and Transfection

A full length cDNA clone was cut with EcoN1 and BstX1 to give a fragment which spanned nucleotides 233 to 1019. This fragment was incorporated into pCMVneo plasmids (Tomasetto C, et al. (1993) J. *Cell Biol.* 122:57-167) and checked for correct orientation of the insert. 5x10^6 MDA-MB435C cells were electroporated at 220V with 10mg of this construct in the presence of 10mg/ml DEAE dextran. Vector alone was used as a negative control. 10^6 cells were plated in five P100 dishes in Alpha -5% FCS. After 14 days of selection in media containing 1 mg/ml G418, the transfected clones were refed with media containing 0.5mg/ml G418 for an additional week. Clones were picked in cloning cylinders, expanded and maintained in Alpha-5% FCS containing 0.5mg/ml G418.

RESULTS

Differential Display

Total RNAs from 21 PT and 21 MT-1 cell lines were compared by differential display. Approximately 100 bands appeared in each lane of each primer pair tested, and on the average 2-3 bands were differentially expressed. One of the bands that was overexpressed in the 21 PT lane (with primer pair OPA 1/T12MG) (280 bp in Figure 1A) was excised from the gel and PCR amplified. The resulting 280 bp PCR product was used to probe a northern blot (Figure 1B). Two bands were detected; a band of 1.7kb which was very high in 21 PT and barely detectable in 21 MT-1, and a band of approximately 1 kb which was equal in both lanes. This mixture of two clones was purified and the clone which hybridized only to the differentially expressed 1.7 kB message was recovered.

Protease M: Sequence Identification

The 0.28 kb insert was used to screen a 76N cDNA library constructed in lZapII. The longest clone isolated was sequenced in its entirety. This clone is 1,526 nt in length and contains 245bp of 5'nt sequences, 732 bp of coding sequences (coding for a postulated protein of 244 aa), and 549 bp of 3'nt sequences. (Figure 2) The presumptive protein coding region begins with an ATG codon, which is in a good Kozak consensus sequence (Kozak M. (1984) *Nucleic Acids Research* 12:857872) CGGCCATGA, and ends with a TGA translation stop codon. The amino terminal portion of the postulated protein has 13 consecutive hydrophobic residues (leu^4 to ala^{16}) which is characteristic of a signal peptide followed by glu^{17}-glu-gln-asn-lys^{21} which resembles a pro-peptide with a potential trypsin susceptible cleavage site after lys^{21}. A potential N-linked glycosylation site is found at asn^{134}-thr-thr^{136} in the 3'nt region, the expected polyadenylation signal AATAAA is found 11 base pairs upstream of the poly A tail at 1,490 bp. Another polyadenylation signal AATAAA was found at 1,095 bp.

The postulated protein sequence, compared to proteins in the database using the blast program, was similar to other proteins of the serine protease family. The postulated sequence was compared to the four most closely related proteins using the pileup program and distances program and displayed by the prettyplot program and was found to be novel. (Figure 3).

Expression of mRNA in mammary and prostate cells

Figure 4A shows the results of northern blots of mammary cell lines and strains. The two normal cells strains shown (76N and 70N) and another normal cell strain 81N (not shown) expressed the 1.7kb Protease M message at low levels. Two primary tumor lines (21 PT and 21 NT) as well as one metastatic line from the same patient (21 MT-2) expressed high levels of message (approximately 20 to 100 fold higher than the normal strains). However, the most metastatic cell line from the same patient (21 MT-1) expressed low levels of RNA (see Figure 1A). One other primary tumor cell line (BT474) and nine other metastatic cell lines (MCF-7, T47D, ZR-75-1, MDA-MB-157, MDA-MB-231, MDA-MB-361, MDA-MB-435, MDA-MB-436, and BT549) had no detectable message. Figure 4B shows northern blots of prostate cell lines. The normal, immortalized cell strains CF3 and CF91 express moderate levels of Protease M mRNA while another normal immortalized strain, MLC expresses just trace amounts. In contrast all three of the tumor cell lines examined (DU145, LNCaP, and PC3) failed to express any Protease M message.

Expression of mRNA in ovarian cell lines and tissue

A series of normal immortalized and primary tumor derived ovarian cell lines were examined for expression of mRNA for Protease M on northern blots. The message was not expressed in any of the five normal immortalized cell lines, but was detected in five of the eight primary tumor cell lines examined (not shown). We also examined the RNA from a series of normal ovarian tissue and biopsies from primary tumors (one of the two northern blots is shown (Figure 5). While mRNA was not expressed in the three normal tissues examined, the six borderline ovarian tumor tissues, and the two metastatic tumors from colon primaries, it was expressed in the primary ovarian tumor tissue in sixteen of the twenty specimens examined.

Expression of Protease M mRNA in normal human tissue

A blot containing 2 mg of polyA$^+$ RNA from eight normal human tissues (Clontech) was examined for expression of Protease M (Figure 6). While the message was not detected in heart, placenta, lung, liver, or skeletal muscle, high levels of message were detected in brain, kidney, and pancreas. The message detected in brain and kidney was 1.7 to 1.8kb, but the message detected in pancreas was only about 1.2kb. A probable explanation for the smaller message in pancreatic RNA would be the use of the alternative polyadenylation signal at 1090 bp noted in Figure 1.

Production of polyclonal antibody and its use to study expression of protein in mammary cell lines and strains

A polyclonal antibody was produced in rabbits against a hydrophilic peptide which was not highly conserved among other serine proteases (^{73}GKHNLRQRESSQEQS^{87}). The western blot (Figure 7) shows that the antibody detects a protein of 37kd in total cell lysates of the normal mammary epithelial cell strain 81N, and in the primary tumor cell line 21NT. No protein is detected in the metastatic breast cell line MDA-MB-435. In other western blots (not shown) the antibody detected a 37kd protein in the normal strains 70N and 76N, as well as the primary tumor cell line 21PT, but not in the metastatic cell lines T47D and MCF-7. Up to one ml of conditioned media from 70N and 21NT was examined in western blots probed with this antibody and no reacting proteins were detected (not shown). This result suggests that the protein is primarily localized intracellularly and not secreted. The protein detected by the antibody is 37kd while the amino acid sequence predicts a protein of about 27kd. The potential glycosylation site at (^{134}asn-thr-thr^{136}) might explain this size discrepancy.

Table 1 shows that the RNA levels for the serine protease are not always correlated with the protein levels. While the primary tumor cell lines (21NT and 21PT) have 20 to 100 times more Protease M mRNA than normal cell strains (70N, 76N, and 81N), the protein detected on westerns is equal to or somewhat lower for the primary tumor cell lines than in the normal cell strains.

The antipeptide polyclonal Protease M antibody has been used successfully in western blots but does not seem to work in cellular immunofluorescence studies in which the antibody gave a high background with MDA-MB-435 cells which do not express the Protease M message.

Production of Recombinant Protein

Extensive efforts were made to produce recombinant protein for further study of the protease. As briefly discussed below, neither production in E. coli as a GST-fusion protein nor in baculovirus as a pure protein were successful in providing more than minimal amounts of the protease. The products which were recovered were used primarily to verify the specificity of the antibody preparations.

In a further effort to obtain recombinant protein, transfectants were produced expressing Protease M in the mammary tumor cell line MDA-MB-435. Transfectants were screened initially for protein production, and as shown below, the results demonstrated that only 5 of the 76 transfectants produced any protein and this was at low levels.

Production of GST fusion protein and assay for protease activity

The expected 52kd GST/Protease M fusion protein was purified and yielded approximately 600 mg of fusion protein per 500 ml culture. We were able to cleave the fusion protein by incubation with thrombin but the Protease M fragment was degraded, even at limiting dilutions, while only the GST portion stayed intact. When we ran the fusion protein on western blots, we needed at least 1mg to get a detectable signal.

Up to 1mg of GST/Protease M fusion protein was run on casein and gelatin zymograms (Heusen C, Dowdle EB. (1980) *Anal. Biochem.* 102:196-202; Wilson MJ, et al. (1993) *Journal of Urology* 149:65 3 -65 8) with no evidence of any protease activity while as little as 0.5 ng of bovine trypsin gave detectable protease activity. 5mg of fusion protein did not cleave the chromogenic trypsin substrate BAEE while 1mg of trypsin gave consistently positive results.

Production of Baculovirus Recombinant Protein

50mg of lysates prepared from sf9 cells infected with an amplified stock of Protease M recombinant baculovirus were run on a western blot and probed with anti Protease M antibody (Figure 7). While no reacting proteins were seen in the lysate from uninfected sf9 cells, a protein of 39kd was detected in lysates of sf9 infected with recombinant baculovirus. Sf9/1G3(Schachter M. (1980) Kallikreins (kininogenases) *Pharmacol. Rev.* 31: 1 -1 7) had approximately 50% infected, enlarged cells while sf9/1G3(Reigman PH, Vlietstra RJ, Suurmeijer L, et al. (1992) *Genomics* 14:6-11.) which was infected with 5 times more virus had nearly 100% infected cells. However, the amount of recombinant protein was quite low and we were unable to detect a band of 39kd on commasie blue stained gels (not shown).

The best purification protocol for purification of recombinant Protease M from lysates was p-aminobenzamidine agarose affinity chromatography followed by concanavalin A agarose chromatography. Using this protocol, recombinant Protease M was purified approximately 80-fold. However, the protein was still only 10% pure (judging from silver-stained gels) and the yield was calculated to be less than 1mg/10^8 cells. Using this data we were able to calculate that 50 mg of lysate contains 15 ng of Protease M or 0.03 % of the total protein. Furthermore, by comparing the amount of the 39kd band determined on silver stained gels of the 80-fold purified Protease M, with western blots of the purified protein, we were able to determine that the antibody can detect 5 ng of Protease M protein as a lower limit. Up to 100mg of lysate or 100 ng of 80-fold purified Protease M (containing approximately 10 ng of Protease M) was run on on gelatin and casein zymograms and failed to detect protease activity (not shown). Up to 0.5 ng of trypsin run in parallel lanes was detected.

MDA-MB435 Transfectants

A pCMV/Neo/Protease M construct as well as neo-vector controls were transfected into MDA-MB435 cells (5 x 10^6 cells for each construct) by electroporation. Of the 10^6 cells which survived the electroporation, approximately 400 colonies (one in 2,500) survived the G418 selection. 80 colonies of protease transfected clones and 20 colonies of vector transfected clones were transferred to 24 well dishes when they were 2 to 3 mm in diameter. The protease transfected cells grew more slowly and had more enlarged, dying cells than the vector controls. Total cell lysates were prepared from the 76 protease transfectants when the cells were approximately 70% confluent. Western blots, prepared from 50mg of the lysate from the 76 transfectants as well as 50mg of lysate from 70N (positive control), were probed with the Protease M antibody. Only 2 of 42 fast growing clones and 2 of 34 slow growing clones expressed any detectable

protein. (data not shown). Furthermore, the level of protein expressed by these positive clones, was, in all cases, considerably less than in 70N cells.

Table 2 shows that Protease M RNA was found in clones expressing protein as well as the majority of those not expressing protein. Thus, in MDA-MB-435 cells there is either inefficient translation of the message, or the protein translated is extremely unstable.

Table 1 Shows the Expression of Protease M mRNA and protein in mammary cells

Cell Line	RNA[1]	Protein[2]
70N	5	100
81N	4	60
16-1-1 (76N/HPV16)	4	64
21NT	85	47
21PT	100	76
MDA.MB435	0	0
T47D	0	0
MCF-7	0	0

[1]RNA values were obtained by running 10mg of total RNA on a northern blot, hybridizing to ^{32}P-labeled Protease M probes and quantitating the resulting autoradiograms. The most intense band was set equal to 100 and the other values normalized accordingly [2] protein values were obtained by running 50mg of total cell lysates on a western blot and probing with the Protease M antibody as described in methods. The 37kd bands on the autoradiograms were quantitated, the most intense band was set equal to 100 and the other values normalized accordingly.

Table 2 Shows the Analysis of Protease M RNA and Protein Expression in MDA-MB-435 Tranfectants

Cell Line	RNA[1]	Protein[2]
70N	12	100
MDA.MB435	0	0
Protease M transfectant #13	4	0
#19	10	0
#42	96	25
#44	61	12
#53	0	0
#58	100	0
#59	6	0
#64	22	0
#65	44	25
#66	55	63
#75	22	0
#86	0	0

[1,2] these values were determined as in the footnote to Table 1.

Table 3.Shows NORTHERN RESULTS WITH OVARIAN CELL LINES AND TISSUES

CELL LINES	FRACTION OF CELLS EXPRESSING
NORMAL	0/5
TUMOR	5/8
TISSUES	
NORMAL	0/2
TUMOR	13/19

Table 4.

CELL LINE	DESCRIPTION	Protease M RNA EXP	6A2 RNA EXP
70N	Normal human mammary epithelial	+	+
76N	normal human mammary epithelial	+	-
81N	normal human mammary epithelial	+	+
21NT	primary breast carcinoma	++ to +++	++ to +++
21PT	primary breast carcinoma	++++	++++
21MT2	metastatic breast carcinoma (pleural effusion)	+ to +++	++ to +++
21MT1	metastatic breast carcinoma (pleural effusion)	+	+
MDA-MB-157	breast medulla carcinoma (pleural effusion)	-	-
MDA-MB	Breast adenocarcinoma (pleural effusion)	-	-
MDA-MB-361	breast adenocarcinoma (brain metastasis)	-	-
MDA-MB-435	breast ductal carcinoma (pleural effusion)	-	-

CELL LINE	DESCRIPTION	Protease M RNA EXP	6A2 RNA EXP
MDA-MB-436	breast adenocarcinoma (pleural effusion)	-	-
BT-474	breast invasive ductal carcinoma (primary)	-	-
BT-549	breast invasive ductal carcinoma (metastasis to lymph nodes)	-	-
HS578T	breast ductal carcinoma (primary)	-	-
MCF7	breast adenocarcinoma (primary	-	-
T-47D	breast ductal carcinoma (pleural effusion)	-	-
ZR-75-1	breast ductal carcinoma (ascitic effusion)	-	-
56NF	normal breast fibroblast	-	-
PC-3 (CRL1435)	prostate adenocarcinoma	-	-
WiDR (CCL218)	colon adenocarcinoma	++	(+)
SW48 (CCL228)	colon adenocarcinoma	+	(+)
MIA Pa-CA-2 (CCL1420)	pancreatic carcinoma	-	-
HuTu 80	duodenal adenocarcinoma	-	-

CELL LINE	DESCRIPTION	Protease M RNA EXP	6A2 RNA EXP
T24	bladder transitional cell carcinoma	-	-
A549(CCL185)	lung carcinoma	-	-
Calu-1	lung epidermoid carcinoma	-	-
Oat 4	lung small cell carcinoma	-	-
G-361	malignant melanoma	-	-
SMKE 30	malignant melanoma	-	-
A2058	malignant melanoma	-	-
SCC-25	tongue squamous cell carcinoma	-	-
RD	rhadomyosarcoma of pelvis	-	-
Kaposi	kaposis sarcoma	-	-
FS3	foreskin fibroblast	-	-
Leukocyte	normal leukocytes	-	-

TABLE 5. SHOWS RNA EXPRESSION IN MAMMARY TISSUE

SAMPLE	TYPE	Protease M	MASPIN	CX26	CX43
81N	N cell strain	++	++++	+++	++++
MDA-MB-435	T cell line	-	-	-	-
CHTN 4253B	CA	+++	+	+/-	+
CHTN 4420A	CA	-	-	+	
CHTN 4782B	CA	+++	+++	++	++
CHTN 5075A	CA	-	-	+	-
CHTN 4253A	NAT	-	-	-	++
CHTN 4303	NAT	++	++	-	+++
CHTN 6281E	NAT	+	+	-	++
CHTN 4728A	RM	-	-	-	++
CHTN 4760C	RM	-	-	-	++
CHTN 5303A	RM	+++	++	-	++
RM (10/30/87)	RM	+	+	-	++
RM-70N	RM	+	+	-	++
RM-70N	RM	+	+	-	++
RM-83N	RM	+	+	-	++

EQUIVALENTS

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: DANA-FARBER CANCER INSTITUTE
(B) STREET: 44 BINNEY STREET
(C) CITY: BOSTON
(D) STATE: MASSACHUSETTS
(E) COUNTRY: US
(F) POSTAL CODE (ZIP): 02115
(G) TELEPHONE:
(H) TELEFAX:

(ii) TITLE OF INVENTION: PROTEASE M, A NOVEL SERINE PROTEASE

(iii) NUMBER OF SEQUENCES: 2

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: LAHIVE & COCKFIELD, LLP
(B) STREET: 28 STATE STREET
(C) CITY: BOSTON
(D) STATE: MASSACHUSETTS
(E) COUNTRY: US
(F) ZIP: 02109-1875

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: PCT/US97/
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 60/025,301
(B) FILING DATE: 13 SEPTEMBER 1996

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: MANDRAGOURAS, AMY E.
(B) REGISTRATION NUMBER: 36,207
(C) REFERENCE/DOCKET NUMBER: DFN-009PC

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (617)227-7400
(B) TELEFAX: (617)742-4214

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1526 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 246..978

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
AGGCGGACAA AGCCCGATTG TTCCTGGGCC CTTTCCCCAT CGCGCCTGGG CCTGCTCCCC    60

AGCCCGGGGC AGGGGCGGGG GCCAGTGTGG TGACACACGC TGTAGCTGTC TCCCCGGCTG   120

GCTGGCTCGC TCTCTCCTGG GGACACAGAG GTCGGCAGGC AGCACACAGA GGGACCTACG   180

GGCAGCTGTT CCTTCCCCCG ACTCAAGAAT CCCCGGAGGC CCGGAGGCCT GCAGCAGGAG   240

CGGCC ATG AAG AAG CTG ATG GTG GTG CTG AGT CTG ATT GCT GCA GCC         287
      Met Lys Lys Leu Met Val Val Leu Ser Leu Ile Ala Ala Ala
        1               5                  10

TGG GCA GAG GAG CAG AAT AAG TTG GTG CAT GGC GGA CCC TGC GAC AAG      335
Trp Ala Glu Glu Gln Asn Lys Leu Val His Gly Gly Pro Cys Asp Lys
 15                  20                  25                  30

ACA TCT CAC CCC TAC CAA GCT GCC CTC TAC ACC TCG GGC CAC TTG CTC      383
Thr Ser His Pro Tyr Gln Ala Ala Leu Tyr Thr Ser Gly His Leu Leu
                35                  40                  45

TGT GGT GGG GTC CTT ATC CAT CCA CTG TGG GTC CTC ACA GCT GCC CAC      431
Cys Gly Gly Val Leu Ile His Pro Leu Trp Val Leu Thr Ala Ala His
            50                  55                  60

TGC AAA AAA CCG AAT CTT CAG GTC TTC CTG GGG AAG CAT AAC CTT CGG      479
Cys Lys Lys Pro Asn Leu Gln Val Phe Leu Gly Lys His Asn Leu Arg
        65                  70                  75

CAA AGG GAG AGT TCC CAG GAG CAG AGT TCT GTT GTC CGG GCT GTG ATC      527
Gln Arg Glu Ser Ser Gln Glu Gln Ser Ser Val Val Arg Ala Val Ile
    80                  85                  90

CAC CCT GAC TAT GAT GCC GCC AGC CAT GAC CAG GAC ATC ATG CTG TTG      575
His Pro Asp Tyr Asp Ala Ala Ser His Asp Gln Asp Ile Met Leu Leu
 95                  100                 105                 110

CGC CTG GCA CGC CCA GCC AAA CTC TCT GAA CTC ATC CAG CCC CTT CCC      623
Arg Leu Ala Arg Pro Ala Lys Leu Ser Glu Leu Ile Gln Pro Leu Pro
                115                 120                 125
```

```
CTG GAG AGG GAC TGC TCA GCC AAC ACC ACC AGC TGC CAC ATC CTG GGC     671
Leu Glu Arg Asp Cys Ser Ala Asn Thr Thr Ser Cys His Ile Leu Gly
            130                 135                 140

TGG GGC AAG ACA GCA GAT GGT GAT TTC CCT GAC ACC ATC CAG TGT GCA     719
Trp Gly Lys Thr Ala Asp Gly Asp Phe Pro Asp Thr Ile Gln Cys Ala
        145                 150                 155

TAC ATC CAC CTG GTG TCC CGT GAG GAG TGT GAG CAT GCC TAC CCT GGC     767
Tyr Ile His Leu Val Ser Arg Glu Glu Cys Glu His Ala Tyr Pro Gly
    160                 165                 170

CAG ATC ACC CAG AAC ATG TTG TGT GCT GGG GAT GAG AAG TAC GGG AAG     815
Gln Ile Thr Gln Asn Met Leu Cys Ala Gly Asp Glu Lys Tyr Gly Lys
175                 180                 185                 190

GAT TCC TGC CAG GGT GAT TCT GGG GGT CCG CTG GTA TGT GGA GAC CAC     863
Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Gly Asp His
                195                 200                 205

CTC CGA GGC CTT GTG TCA TGG GGT AAC ATC CCC TGT GGA TCA AAG GAG     911
Leu Arg Gly Leu Val Ser Trp Gly Asn Ile Pro Cys Gly Ser Lys Glu
            210                 215                 220

AAG CCA GGA GTC TAC ACC AAC GTC TGC AGA TAC ACG AAC TGG ATC CAA     959
Lys Pro Gly Val Tyr Thr Asn Val Cys Arg Tyr Thr Asn Trp Ile Gln
        225                 230                 235

AAA ACC ATT CAG GCC AAG T GACCCTGACA TGTGACATCT ACCTCCCGAC         1008
Lys Thr Ile Gln Ala Lys
    240

CTACCACCCC ACTGGCTGGT TCCAGAACGT CTCTCACCTA GACCTTGCCT CCCCTCCTCT  1068

CCTGCCCAGC TCTGACCCTG ATGCTTAATA AACGCAGCGA CGTGAGGGTC CTGATTCTCC  1128

CTGGTTTTAC CCCAGCTCCA TCCTTGCATC ACTGGGGAGG ACGTGATGAG TGAGGACTTG  1188

GGTCCTCGGT CTTACCCCCA CCACTAAGAG AATACAGGAA AATCCCTTCT AGGCATCTCC  1248

TCTCCCCAAC CCTTCCACAC GTTTGATTTC TTCCTGCAGA GGCCCAGCCA CGTGTCTGGA  1308

ATCCCAGCTC CGCTGCTTAC TGTCGGTGTC CCCTTGGGAT GTACCTTTCT TCACTGCAGA  1368

TTTCTCACCT GTAAGATGAA GATAAGGATG ATACAGTCTC CATCAGGCAG TGGCTGTTGG  1428

AAAGATTTAA GATTTCACAC CTATGACATA CATGGGATAG CACCTGGGCC GCCATGCACT  1488

CAATAAAGAA TGTATTTTAA AAAAAAAAAA AAAAAAAA                          1526
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 244 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Lys Lys Leu Met Val Val Leu Ser Leu Ile Ala Ala Ala Trp Ala
 1               5                  10                  15

Glu Glu Gln Asn Lys Leu Val His Gly Gly Pro Cys Asp Lys Thr Ser
            20                  25                  30

His Pro Tyr Gln Ala Ala Leu Tyr Thr Ser Gly His Leu Leu Cys Gly
        35                  40                  45

Gly Val Leu Ile His Pro Leu Trp Val Leu Thr Ala Ala His Cys Lys
    50                  55                  60

Lys Pro Asn Leu Gln Val Phe Leu Gly Lys His Asn Leu Arg Gln Arg
65                  70                  75                  80

Glu Ser Ser Gln Glu Gln Ser Ser Val Val Arg Ala Val Ile His Pro
                85                  90                  95

Asp Tyr Asp Ala Ala Ser His Asp Gln Asp Ile Met Leu Leu Arg Leu
            100                 105                 110

Ala Arg Pro Ala Lys Leu Ser Glu Leu Ile Gln Pro Leu Pro Leu Glu
        115                 120                 125

Arg Asp Cys Ser Ala Asn Thr Thr Ser Cys His Ile Leu Gly Trp Gly
    130                 135                 140

Lys Thr Ala Asp Gly Asp Phe Pro Asp Thr Ile Gln Cys Ala Tyr Ile
145                 150                 155                 160

His Leu Val Ser Arg Glu Glu Cys Glu His Ala Tyr Pro Gly Gln Ile
                165                 170                 175

Thr Gln Asn Met Leu Cys Ala Gly Asp Glu Lys Tyr Gly Lys Asp Ser
            180                 185                 190

Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Gly Asp His Leu Arg
        195                 200                 205

Gly Leu Val Ser Trp Gly Asn Ile Pro Cys Gly Ser Lys Glu Lys Pro
    210                 215                 220

Gly Val Tyr Thr Asn Val Cys Arg Tyr Thr Asn Trp Ile Gln Lys Thr
225                 230                 235                 240

Ile Gln Ala Lys
```

What is claimed is:

1. An isolated nucleic acid molecule comprising a nucleotide sequence encoding Protease M or a biologically active portion thereof.

2. An isolated nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1.

3. An isolated nucleic acid molecule at least 15 nucleotides in length which hybridizes under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1.

4. The isolated nucleic acid molecule of claim 1, comprising the coding region of the nucleotide sequence of SEQ ID NO: 1.

5. An isolated nucleic acid molecule comprising a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least 60 % homologous to the amino acid sequence of SEQ ID NO: 2.

6. The isolated nucleic acid molecule of claim 5, wherein the protein comprises an amino acid sequence at least 70 % homologous to the amino acid sequence of SEQ ID NO: 2

7. The isolated nucleic acid molecule of claim 5, wherein the protein comprises an amino acid sequence at least 80 % homologous to the amino acid sequence of SEQ ID NO: 2.

8. The isolated nucleic acid molecule of claim 5, wherein the protein comprises an amino acid sequence at least 90 % homologous to the amino acid sequence of SEQ ID NO: 2.

9. An isolated nucleic acid molecule encoding the amino acid sequence of SEQ ID NO: 2.

10. An isolated nucleic acid molecule encoding a Protease M fusion protein.

11. An isolated nucleic acid molecule which is antisense to the nucleic acid molecule of claim 1.

12. The isolated nucleic acid molecule of claim 1 which is antisense to a coding region of the coding strand of the nucleotide sequence of SEQ ID NO: 1.

13. The isolated nucleic acid molecule of claim 1 which is antisense to a noncoding region of the nucleotide sequence of SEQ ID NO: 1.

14. The isolated nucleic acid molecule of claim 1 isoloated using at least a portion of the nucleotide sequence of SEQ ID NO:1 as a probe or a primer.

15. A vector comprising a nucleotide sequence encoding Protease M.

16. The vector of claim 15, which is a recombinant expression vector.

17. The vector of claim 16, which encodes a protein comprising the amino acid sequence of SEQ ID NO: 2.

18. The vector of claim 15, which comprises the coding region of the nucleotide sequence of SEQ ID NO: 1.

19. A host cell containing the vector of claim 17.

20. A host cell containing the recombinant expression vector of claim 18.

21. A method for producing Protease M comprising culturing the host cell of claim 19 in a suitable medium until Protease M is produced.

22. The method of claim 21, further comprising isolating Protease M from the medium or the host cell.

23. An isolated Protease M protein or a biologically active portion thereof.

24. An isolated Protease M protein, wherein said protein is encoded by the nucleic acid shown in SEQ ID No:1.

25. An isolated protein which comprises an amino acid sequence at least 60 % homologous to the amino acid sequence of SEQ ID NO: 2.

26. An isolated protein which comprises an amino acid sequence at least 70 % homologous to the amino acid sequence of SEQ ID NO: 2.

27. An isolated protein which comprises an amino acid sequence at least 80 % homologous to the amino acid sequence of SEQ ID NO: 2.

28. An isolated protein which comprises an amino acid sequence at least 90 % homologous to the amino acid sequence of SEQ ID NO: 2.

29. An isolated protein comprising amino acids 22-244 of SEQ ID NO: 2.

30. A pharmaceutical composition comprising the protein of SEQ ID No:2 or biologically active portion thereof and a pharmaceutically acceptable carrier.

31. A fusion protein comprising a Protease M polypeptide operatively linked to a non-protease M polypeptide.

32. An antigenic peptide of Protease M comprising at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO: 2, the peptide comprising an epitope of Protease M such that an antibody raised against the peptide forms a specific immune complex with Protease M.

33. An antibody that specifically binds Protease M.

34. The antibody of claim 33, which is monoclonal.

35. The antibody of claim 34, which is coupled to a detectable substance.

36. A pharmaceutical composition comprising the antibody of claim 34 and a pharmaceutically acceptable carrier.

37. A nonhuman transgenic animal which contains cells carrying a transgene encoding Protease M.

38. A nonhuman homologous recombinant animal which contains cells having an altered Protease M gene.

39. A method for detecting the presence of Protease M in a biological sample comprising contacting a biological sample with an agent capable of detecting Protease M protein or nucleic acid.

40. The method of claim 39, wherein the agent is a labeled or labelable nucleic acid probe capable of hybridizing to Protease M nucleic acid.

41. The method of claim 40, wherein the agent is a labeled or labelable antibody capable of specifically binding to Protease M protein.

42. The method of claim 40, wherein the biological sample is a tumor sample.

43. The method of claim 40, wherein the tumor sample is a mammary tumor sample.

44. A kit for detecting the presence of protease M in a biological sample comprising a labeled or labelable agent capable of detecting protease M protein or nucleic acid in a biological sample; means for determining the degree of binding to the sample; and means for comparing the amount of amount of binding to the sample with a standard.

45. The kit of claim 44, wherein the agent is a nucleic acid probe capable of hybridizing to protease M nucleic acid.

46. The kit of claim 44, wherein the agent is an antibody capable of specifically binding to protease M protein.

47. A method comprising contacting a cell with an agent that modulates protease M serine proteinase activity associated with the cell.

48. The method of claim 47, wherein the agent stimulates the protease M cysteine proteinase inhibitory activity associated with the cell.

49. The method of claim 47, wherein the agent inhibits the protease M serine proteinase activity associated with the cell.

50. The method of claim 48, wherein the agent is an active protease M protein.

51. The method of claim 48, wherein the agent is a nucleic acid encoding protease M that has been introduced into the cell.

52. The method of claim 49, wherein the agent is an antisense protease M nucleic acid molecule.

55. The method of claim 49, wherein the agent is an antibody that specifically binds to protease M.

56. The method of claim 47, wherein the cell is present within a subject and the agent is administered to the subject.

57. A method for inhibiting development or progression of a metastatic phenotype in a tumor cell comprising contacting the tumor cell with an agent which modulates the amount of or activity of protease M in or around the tumor cell.

58. The method of claim 57, wherein the agent is protease M.

59. The method of claim 57, wherein the agent is a nucleic acid encoding protease M that has been introduced into the tumor cell.

60. The method of claim 57, wherein the agent is a nucleic acid antisense to protease M that has been introduced into the tumor cell.

61. The method of claim 57, wherein the tumor cell is a mammary tumor cell.

62. A method for identifying a modulator of the serine protease activity of protease M, comprising

incubating protease M, a serine protease, a substrate for the serine protease and a test substance under conditions suitable for the serine protease to cleave the substrate;

measuring the cleavage of the substrate;

comparing the amount of cleavage of the substrate in the presence of the test substance to the amount of cleavage of the substrate in the absence of the test substance; and

identifying the test substance as a modulator of the serine protease inhibitory activity of protease M.

63. A method for identifying a modulator of protease M expression, comprising

contacting a cell with a test substance;

determining the level of expression of protease M mRNA or protein in the cell;

comparing the level of expression of protease M mRNA or protein in the cell in the presence of the test substance to level of expression of protease M mRNA or protein in the cell in the absence of the test substance; and

identifying the test substance as a modulator of protease M expression.

FIG.1A
DD Gel

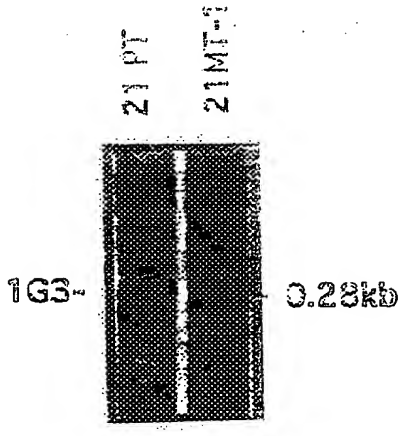
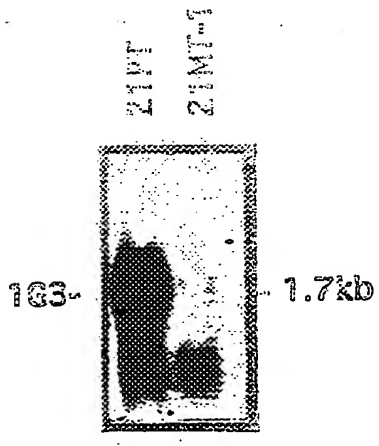


FIG.1B
Northern



PROTEASE M SEQUENCE

```
1     AGGCGGACAAAGCCCGATTGTTCCTGGGCCCTTTCCCCATCGCGCCTGGGCCTGCTCCCCAGCCCGG      67
8     GGCAGGGGCGGGGGCCAGTGTGGTGACACACGCTGTAGCTGTCTCCCCGGCTGGCTGGCTCGCTCTC      134
5     TCCTGGGGACACAGAGGTCGGCAGGCAGCACACAGAGGGACCTACGGGCAGCTGTTCCTTCCCCCGA      201
2     CTCAAGAATCCCCGGAGGCCCGGAGGCCTGCAGCAGGAGCGGCC  245

6     ATG AAG AAG CTG ATG GTG GTG CTG AGT CTG ATT GCT GCA GCC TGG GCA↓GAG      296
1     Met Lys Lys Leu Met Val Val Leu Ser Leu Ile Ala Ala Ala Trp Ala Glu      17

97    GAG CAG AAT AAG↓TTG GTG CAT GGC GGA CCC TGC GAC AAG ACA TCT CAC CCC      347
18    Glu Gln Asn Lys Leu Val His Gly Gly Pro Cys Asp Lys Thr Ser His Pro      34

48    TAC CAA GCT GCC CTC TAC ACC TCG GGC CAC TTG CTC TGT GGT GGG GTC CTT      398
35    Tyr Gln Ala Ala Leu Tyr Thr Ser Gly His Leu Leu Cys Gly Gly Val Leu      51

99    ATC CAT CCA CTG TGG GTC CTC ACA GCT GCC(CAC)TGC AAA AAA CCG AAT CTT      449
52    Ile His Pro Leu Trp Val Leu Thr Ala Ala(His)Cys Lys Lys Pro Asn Leu      68

50    CAG GTC TTC CTG GGG AAG CAT AAC CTT CGG CAA AGG GAG AGT TCC CAG GAG      500
69    Gln Val Phe Leu Gly Lys His Asn Leu Arg Gln Arg Glu Ser Ser Gln Glu      85

01    CAG AGT TCT GTT GTC CGG GCT GTG ATC CAC CCT GAC TAT GAT GCC GCC AGC      551
86    Gln Ser Ser Val Val Arg Ala Val Ile His Pro Asp Tyr Asp Ala Ala Ser      102

52    CAT GAC CAG(GAC)ATC ATG CTG TTG CGC CTG GCA CGC CCA GCC AAA CTC TCT      602
03    His Asp Gln(Asp)Ile Met Leu Leu Arg Leu Ala Arg Pro Ala Lys Leu Ser      119

03    GAA CTC ATC CAG CCC CTT CCC CTG GAG AGG GAC TGC TCA GCC[AAC ACC ACC]      653
20    Glu Leu Ile Gln Pro Leu Pro Leu Glu Arg Asp Cys Ser Ala[Asn Thr Thr]      136

54    AGC TGC CAC ATC CTG GGC TGG GGC AAG ACA GCA GAT GGT GAT TTC CCT GAC      704
37    Ser Cys His Ile Leu Gly Trp Gly Lys Thr Ala Asp Gly Asp Phe Pro Asp      153

705   ACC ATC CAG TGT GCA TAC ATC CAC CTG GTG TCC CGT GAG GAG TGT GAG CAT      755
154   Thr Ile Gln Cys Ala Tyr Ile His Leu Val Ser Arg Glu Glu Cys Glu His      170

756   GCC TAC CCT GGC CAG ATC ACC CAG AAC ATG TTG TGT GCT GGG GAT GAG AAG      806
171   Ala Tyr Pro Gly Gln Ile Thr Gln Asn Met Leu Cys Ala Gly Asp Glu Lys      187

807   TAC GGG AAG[GAT]TCC TGC CAG GGT GAT(TCT)GGG GGT CCG CTG GTA TGT GGA      857
188   Tyr Gly Lys[Asp]Ser Cys Gln Gly Asp(Ser)Gly Gly Pro Leu Val Cys Gly      204

858   GAC CAC CTC CGA GGC CTT GTG TCA TGG GGT AAC ATC CCC TGT GGA TCA AAG      908
205   Asp His Leu Arg Gly Leu Val Ser Trp Gly Asn Ile Pro Cys Gly Ser Lys      221

909   GAG AAG CCA GGA GTC TAC ACC AAC GTC TGC AGA TAC ACG AAC TGG ATC CAA      959
222   Glu Lys Pro Gly Val Tyr Thr Asn Val Cys Arg Tyr Thr Asn Trp Ile Gln      238

960   AAA ACC ATT CAG GCC AAG   977
239   Lys Thr Ile Gln Ala Lys   244

978   TGACCCTGACATGTGACATCTACCTCCCGACCTACCACCCCACTGGCTGGTTCCAGAACGTCTCTCA      1044
1045  CCTAGACCTTGCCTCCCCTCCTCTCCTGCCCAGCTCTGACCCTGATGCTTAATAAACGCAGCGACGT      1111
1112  GAGGGTCCTGATTCTCCCTGGTTTTACCCCAGCTCCATCCTTGCATCACTGGGGAGGACGTGATGAG      1178
1179  TGAGGACTTGGGTCCTCGGTCTTACCCCCACCACTAAGAGAATACAGGAAAATCCCTTCTAGGCATC      1245
1246  TCCTCTCCCCAACCCTTCCACACGTTTGATTTCTTCCTGCAGAGGCCCAGCCACGTGTCTGGAATCC      1312
1313  CAGCTCCGCTGCTTACTGTCGGTGTCCCCTTGGGATGTACCTTTCTTCACTGCAGATTTCTCACCTG      1379
1380  TAAGATGAAGATAAGGATGATACAGTCTCCATCAGGCAGTGGCTGTTGGAAAGATTTAAGATTTCAC      1446
1447  ACCTATGACATACATGGGATAGCACCTGGGCCGCCATGCACTCAATAAAGAATGTATTTTAAAAAAA      1513
1514  AAAAAAAAAAAA  1526
```

Figure 2

ALIGNMENT OF PROTEASE M WITH CLOSELY RELATED MEMBERS OF SERINE PROTEASE FAMILY

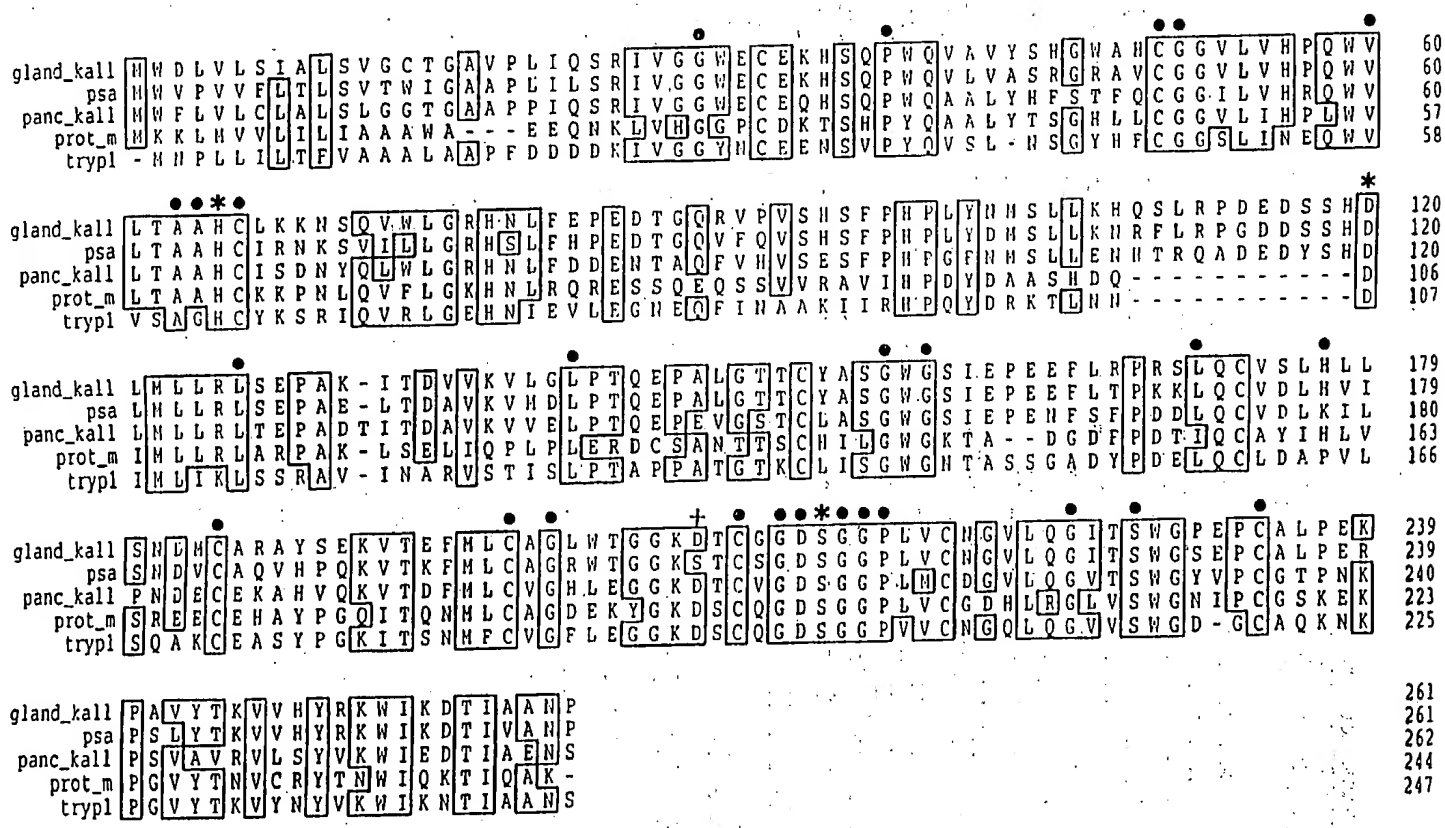


Figure 3

FIG. 4A

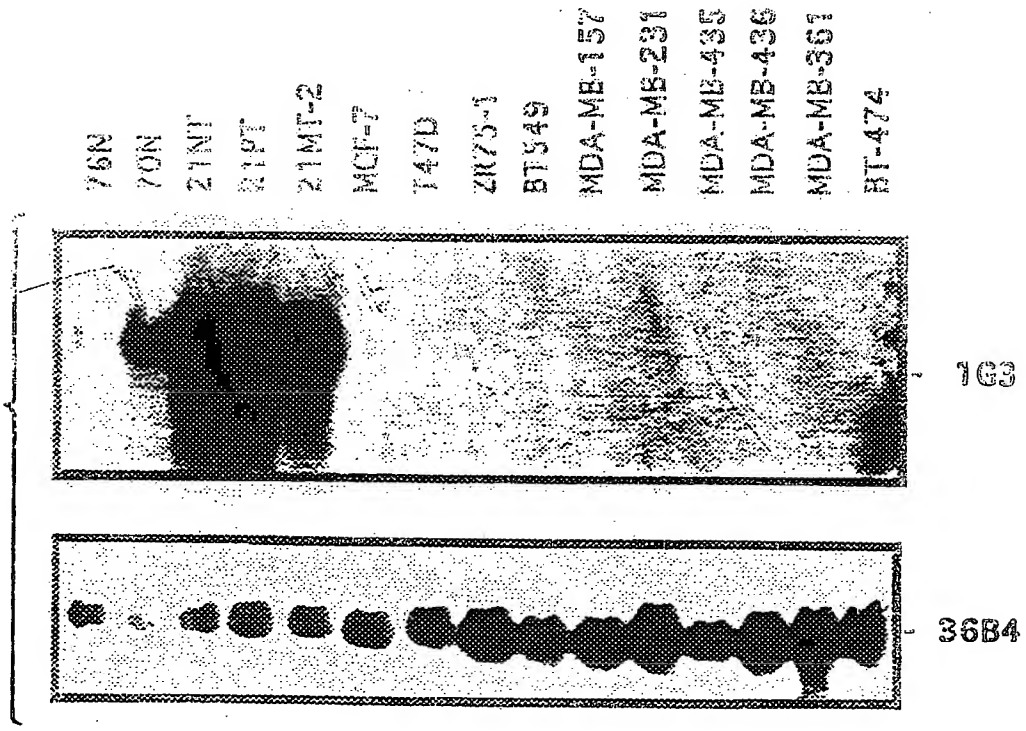

76N
70N
21NT
21PT
21MT-2
MCF-7
T47D
ZR75-1
BT549
MDA-MB-157
MDA-MB-231
MDA-MB-435
MDA-MB-436
MDA-MB-361
BT-474
1G3
36B4

FIG.4B

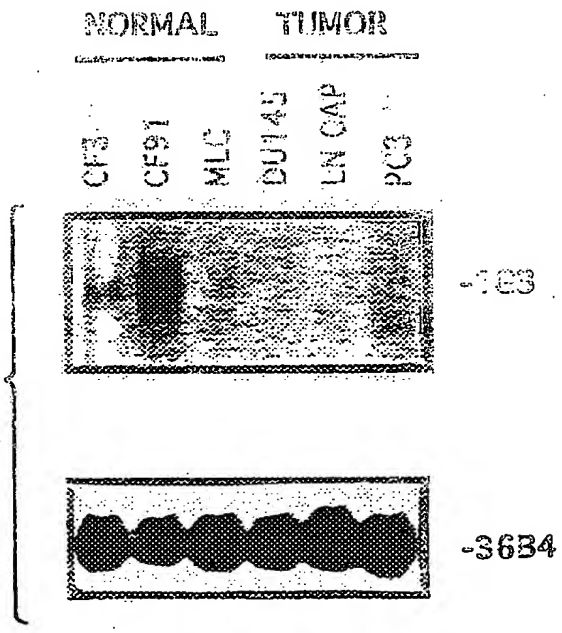


FIG.5

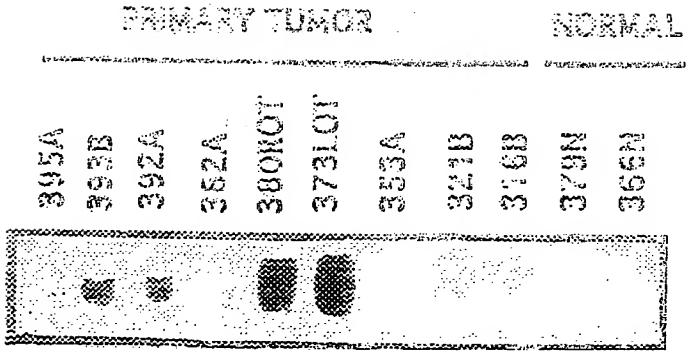

PRIMARY TUMOR
NORMAL
395A
393B
392A
382A
380ROT
373LOT
323A
321B
316B
379N
365N

FIG. 6

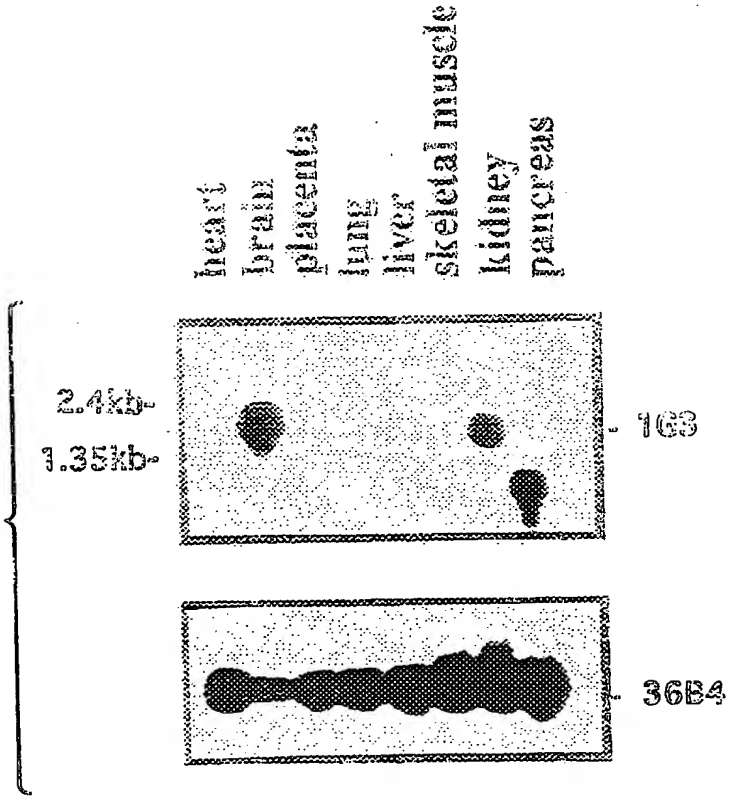
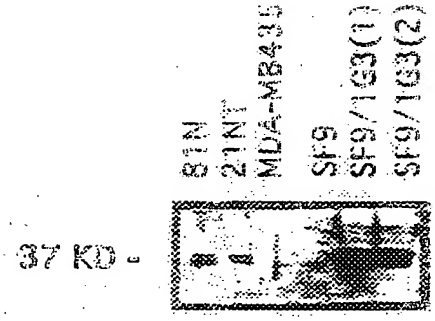

heart
brain
placenta
lung
liver
skeletal muscle
kidney
pancreas
2.4Kb
1.35Kb
1G8
36B4

FIG.7



A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/57 C12N15/62 C12N9/64 C12N15/11 C12N5/10
A61K38/48 C07K16/40 A61K39/395 A01K67/027 C12N15/00
C12Q1/37 C12Q1/68 G01N33/574 A61K31/70 A61K48/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 576 152 A (LILLY CO ELI) 29 December 1993	1-9, 14-29, 32-35, 39-41, 44-46
Y	see the whole document	10-13, 30,31, 36-38, 42,43, 47-63

Y	VIHINEN M: "Modeling of prostate specific antigen and human glandular kallikrein structures." BIOCHEM BIOPHYS RES COMMUN, NOV 15 1994, 204 (3) P1251-6, UNITED STATES, XP002060074 see the whole document	1-9, 14-36, 39-63

	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 March 1998	07. 04. 98

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk	

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	HAGEN HE ET AL: "Simulium damnosum s.l.: identification of inducible serine proteases following an Onchocerca infection by differential display reverse transcription PCR." EXP PARASITOL, NOV 1995, 81 (3) P249-54, UNITED STATES, XP002060075 see the whole document	1-9, 14-29, 32-35, 39-46
Y	EP 0 652 014 A (NAT INST IMMUNOLOGY) 10 May 1995 see claims 1-10	30,36, 47-63
Y	WO 96 26280 A (BASF AG ;KAMENS JOANNE (US); ALLEN HAMISH (US); PASKIND MICHAEL (U) 29 August 1996 see abstract; claims 23,24,28	10-14,31
Y	WO 90 05188 A (PHARMACEUTICAL PROTEINS LTD) 17 May 1990 see abstract; claims 14-17	10,31, 37,38
P,X	ANISOWICZ A ET AL: "A novel protease homolog differentially expressed in breast and ovarian cancer" MOLECULAR MEDICINE (CAMBRIDGE), 2 (5).30 SEP 1996. 624-636., XP002060076 see the whole document	1-9, 14-29, 32-35, 39-41, 44-46

INTERNATIONAL SEARCH REPORT

Int. national application No.
PCT/US 97/16175

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

see FURTHER INFORMATION sheet PCT/ISA/210

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Remark : Although claim 56 and claims 47-55, 57-61 as far as they concern an in vivo method ,are directed to a method of treatment of the human/animal body , the search has been carried out and based on the alleged effects of the compound/composition.

Information on patent family members

PCT/US 97/16175

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0576152 A	29-12-93	AU 3986493 A	02-12-93
		BR 9302075 A	30-11-93
		CA 2096911 A	29-11-93
		CZ 9300982 A	16-02-94
		HU 69612 A	28-09-95
		JP 6062855 A	08-03-94
		MX 9303082 A	30-06-94
		NO 931889 A	29-11-93
		PL 299053 A	21-02-94
EP 0652014 A	10-05-95	NONE	
WO 9626280 A	29-08-96	NONE	
WO 9005188 A	17-05-90	AT 158817 T	15-10-97
		AU 628101 B	10-09-92
		AU 4494389 A	28-05-90
		DE 68928363 D	06-11-97
		EP 0396699 A	14-11-90
		JP 3505674 T	12-12-91
		US 5650503 A	22-07-97